CUK-H03045

1916 75th 1991
ANNIVERSARY

THE ACCOUNTING REVIEW

15

VOLUME 66, No. 1 JANUARY 1991

TERRY SHEVLIN
The Valuation of R&D Firms with R&D Limited Partnerships

A FORUM ON ACCOUNTING DISCLOSURES AND STOCK PRICES

SIVA SWAMINATHAN
The Impact of SEC Mandated Segment Data on Price Variability and Divergence of Beliefs

EARL K. STICE
The Market Reaction to 10-K and 10-Q Filings and to Subsequent *The Wall Street Journal* Earnings Announcements

MARY E. BARTH, WILLIAM H. BEAVER, and CHRISTOPHER H. STINSON
Supplemental Data and the Structure of Thrift Share Prices

MICHAEL J. BRENNAN
A Perspective on Accounting and Stock Prices

VERONIQUE FRUCOT and WINSTON T. SHEARON
Budgetary Participation, Locus of Control, and Mexican Managerial Performance and Job Satisfaction

CLIFTON E. BROWN and IRA SOLOMON
Configural Information Processing in Auditing: The Role of Domain-Specific Knowledge

RAMJI BALAKRISHNAN
Information Acquisition and Resource Allocation Decisions

SHORTER ARTICLES

ROBERT M. HALPERIN and BIN SRINIDHI
U.S. Income Tax Transfer-Pricing Rules and Resource Allocation: The Case of Decentralized Multinational Firms

SHIING-WU WANG
The Relation Between Firm Size and Effective Tax Rates: A Test of Firms' Political Success

MARY STONE and JOHN RASP
Tradeoffs in the Choice Between Logit and OLS for Accounting Choice Studies

EDITOR'S ANNUAL REPORT

A QUARTERLY JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

AMERICAN ACCOUNTING ASSOCIATION

The American Accounting Association was founded in 1916 as the American Association of University Instructors in Accounting. Its present name was adopted in 1936. The Association is a voluntary organization of persons interested in accounting education and research. Full membership entitles one to receive *The Accounting Review*, *Accounting Education News*, *Accounting Horizons*, and *Issues in Accounting Education*, and to purchase Association publications at a discount.

The members of the 1990–91 Executive Committee of the Association are as follows:

President	Alvin A. Arens, Michigan State University
President-Elect	Arthur Wyatt, Arthur Andersen & Co.
Vice President	Wanda A. Wallace, Texas A&M University
Vice President	Mark A. Wolfson, Stanford University
Vice President	David A. Wilson, Ernst & Young
Secretary-Treasurer	Joseph J. Schultz, Jr., Arizona State University
Director of Education	Corine T. Norgaard, University of Connecticut
Director of Research	Nicholas Dopuch, Washington University
Director of Publications	Daniel W. Collins, University of Iowa
Past President	John K. Simmons, University of Florida

The full-time Executive Director of the Association is Paul L. Gerhardt.

Membership in the Association is available at the following annual rates:

Full membership, US and Canada:	US $65.00
Full membership, other countries:	US $35.00
Associate membership:	US $12.00 for students residing in the United States and Canada

657.05
Ac 57 R

In addition, full members may belong to one or more Sections: Auditing; Information Systems/Management Advisory Services; Two-Year Colleges; International Accounting; Public Interest; Government and Nonprofit; Taxation; Administrators of Accounting Programs; Accounting, Behavior and Organizations; and Management Accounting. Each Section assesses its own annual dues which are collected by the Association.

The Accounting Review is published quarterly by the American Accounting Association—January, April, July, and October: printed by George Banta Company, Inc., Menasha, Wisconsin. Subscription price is $90.00 per year.

Second class postage paid at Sarasota, Florida and at additional mailing offices.

Copyright © American Accounting Association. All rights reserved. Printed in the United States of America. ISSN 0001-4826.

Editorial correspondence and all manuscripts should be sent to Professor A. Rashad Abdel-khalik, Editor, *The Accounting Review*, Fisher School of Accounting, College of Business Administration, University of Florida, Gainesville, FL 32611.

The submission fee of $50.00 for members of the AAA or $100.00 for nonmembers must accompany manuscripts when submitted. Books for review should be sent to Professor Shane R. Moriarity, School of Accounting, University of Oklahoma, Norman, OK 73019. Request for reprint permission, subscriptions, advertisements, and dues should be sent to Mr. Paul Gerhardt, American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233. Notice of change of address should be mailed at least eight weeks in advance to ensure undelayed receipt of *The Accounting Review*.

P 8120

THE ACCOUNTING REVIEW

A QUARTERLY JOURNAL OF THE

AMERICAN ACCOUNTING ASSOCIATION

Editor
A. Rashad Abdel-khalik

Associate Editors
Robert H. Ashton
Andrew D. Bailey, Jr.
Bala V. Balachandran
Victor L. Bernard
Ronald W. Hilton
John S. Hughes
Richard A. Lambert
James C. McKeown
Ira Solomon
Shyam Sunder

657.05
Ac57R

⑮

VOL. 66 JANUARY 1991 No. 1

TERRY SHEVLIN
The Valuation of R&D Firms with R&D Limited Partnerships 1

A FORUM ON ACCOUNTING DISCLOSURES AND STOCK PRICES

SIVA SWAMINATHAN
The Impact of SEC Mandated Segment Data on Price Variability and Divergence of Beliefs 23

EARL K. STICE
The Market Reaction to 10-K and 10-Q Filings and to Subsequent *The Wall Street Journal* Earnings Announcements 42

MARY E. BARTH, WILLIAM H. BEAVER, and CHRISTOPHER H. STINSON
Supplemental Data and the Structure of Thrift Share Prices 56

MICHAEL J. BRENNAN
A Perspective on Accounting and Stock Prices 67

VERONIQUE FRUCOT and WINSTON T. SHEARON
Budgetary Participation, Locus of Control, and Mexican Managerial Performance and Job Satisfaction 80

CLIFTON E. BROWN and IRA SOLOMON
Configural Information Processing in Auditing: The Role of Domain-Specific Knowledge 100

RAMJI BALAKRISHNAN
Information Acquisition and Resource Allocation Decisions 120

SHORTER ARTICLES

ROBERT M. HALPERIN and BIN SRINIDHI
U.S. Income Tax Transfer-Pricing Rules and Resource Allocation: The Case of Decentralized Multinational Firms 141

SHIING-WU WANG
The Relation Between Firm Size and Effective Tax Rates: A Test of Firms' Political Success 158

MARY STONE and JOHN RASP
Tradeoffs in the Choice Between Logit and OLS for Accounting Choice Studies 170

EDITOR'S ANNUAL REPORT 188

Contents continued on Overleaf

AMERICAN ACCOUNTING ASSOCIATION
Committees 1990–1991 .... **190**

BOOK REVIEWS, Shane Moriarity, Editor

Blazek, ***Tax and Financial Planning for Tax-Exempt Organizations: Forms, Checklists, Procedures*** .... DARLENE K. AHNBERG **202**
Bromwich and Bhimani, ***Management Accounting: Evolution not Revolution*** .... JAMES A. HENDRICKS **203**
Monden and Sakurai, Editors, ***Japanese Management Accounting: A World Class Approach to Profit Management*** .... KAY M. POSTON **204**
Ruch, ***International Handbook of Corporate Communication*** .... V. K. ZIMMERMAN **205**
Samuelson, ***Models of Accounting Information Systems: The Swedish Case*** .... GRAHAM GAL **206**
Stalk and Hout, ***Competing Against Time—How Time-Based Competition is Reshaping Global Markets*** .... ALFRED J. NANNI, JR. **207**
Thierauf, ***Expert Systems in Finance and Accounting*** .... TERRY L. CAMPBELL **209**
List of Books Received **209**

CONTENTS OF ACCOUNTING RESEARCH JOURNALS **215**

INFORMATION FOR AUTHORS

EDITORIAL POLICY

According to the policies set by the Publications Committee (which were endorsed by the Executive Committee and were published in the *Accounting Education News*, June 1987), *The Accounting Review* "should be viewed as the premier journal for publishing articles reporting the results of accounting research and explaining and illustrating related research methodology. The scope of acceptable articles should embrace any research methodology and any accounting-related subject, as long as the articles meet the standards established for publication in the journal. . . . No special sections should be necessary. The primary, but not exclusive, audience should be—as it is now—academicians, graduate students, and others interested in accounting research."

The primary criterion for publication in *The Accounting Review* is the significance of the contribution an article makes to the literature.

The efficiency and effectiveness of the editorial review process is critically dependent upon the actions of both authors submitting papers and the reviewers. Authors accept the responsibility of preparing research papers at a level suitable for evaluation by independent reviewers. Such preparation, therefore, should include subjecting the manuscript to critique by colleagues and others and revising it accordingly prior to submission. The review process is not to be used as a means of obtaining feedback at early stages of developing the research.

Reviewers and associate editors are responsible for providing critically constructive and prompt evaluations of submitted research papers based on the significance of their contribution and on the rigor of analysis and presentation. Associate Editors also make editorial recommendations to the editor.

(Continued following Book Reviews under Manuscript Preparation and Style)

PAST EDITORS

1926–1929	William A. Paton
1929–1943	Eric L. Kohler
1944–1947	A. C. Littleton
1948–1949	Robert L. Dixon
1950–1959	Frank P. Smith
1960–1962	Robert K. Mautz
1963–1964	Lawrence L. Vance
1965–1967	Wendel Trumbull
1968–1970	Charles H. Griffin
1971–1972	Eldon S. Hendriksen
1973–1975	Thomas F. Keller
1976–1978	Don T. DeCoster
1978–1983	Stephen A. Zeff
1983–1987	Gary L. Sundem
1987–1989	William R. Kinney, Jr.

Editorial Staff of *The Accounting Review*
(1990–91)

MANAGING EDITOR AND EDITOR

A. Rashad Abdel-khalik, University of Florida

ASSOCIATE EDITORS

Robert H. Ashton, Duke University
Andrew D. Bailey, University of Arizona
Bala V. Balachandran, Northwestern University
Victor L. Bernard, University of Michigan
Ronald W. Hilton, Cornell University
John S. Hughes, University of Minnesota
Richard A. Lambert, University of Pennsylvania
James C. McKeown, Pennsylvania State University
Ira Solomon, University of Illinois at Urbana-Champaign
Shyam Sunder, Carnegie Mellon University

Book Review Editor: **Shane Moriarity,** University of Oklahoma

Copy and Production: **Patricia M. Calomeris**

EDITORIAL ADVISORY AND REVIEW BOARD

Matthew Anderson, Michigan State University
Urton L. Anderson, University of Texas at Austin
Anthony A. Atkinson, University of Waterloo
Frances L. Ayres, University of Oklahoma
Rajiv D. Banker, University of Minnesota
Ravi Bhushan, Massachusetts Institute of Technology
Gary C. Biddle, University of Washington
James R. Boatsman, Arizona State University
Peter Brownell, Macquarie University
John Christensen, Odense Universtet
Julie H. Collins, University of North Carolina–Chapel Hill
Barry E. Cushing, Pennsylvania State University
Srikant Datar, Stanford University
Joel S. Demski, Yale University
John W. Dickhaut, University of Minnesota
John H. Evans, University of Pittsburgh
Haim Falk, McMaster University
John E. Fellingham, University of Iowa/Yale University
Michael Gibbins, University of Alberta
Dan Givoly, Tel-Aviv University and Northwestern University
Robert M. Halperin, New York University
James V. Hansen, Brigham Young University
Trevor S. Harris, Columbia University
Paul M. Healy, Massachusetts Institute of Technology
Robert W. Holthausen, University of Pennsylvania
Keith Houghton, University of Melbourne
Herbert G. Hunt, III, University of Vermont
Rhoda C. Icerman, Florida State University
Prem C. Jain, University of Pennsylvania
Chandra Kanodia, University of Minnesota
Lauren Kelly, University of Washington
April Klein, New York University
W. Robert Knechel, University of Florida
S. P. Kothari, University of Rochester
Young K. Kwon, University of Illinois at Urbana-Champaign
David Larcker, University of Pennsylvania
Baruch Lev, University of California at Berkeley
Barry L. Lewis, University of Colorado at Boulder
Robert Libby, Cornell University
Marlys G. Lipe, University of Michigan
Martin P. Loeb, University of Maryland
Russell Lundholm, Stanford University
Robert P. Magee, Northwestern University
Ajay Maindiratta, New York University
Kathleen T. McGaharan, Columbia University
Barbara D. Merino, University of North Texas
Jane F. Mutchler, University of Arizona
Nandu Nagarajan, University of Pittsburgh
Eric W. Noreen, University of Washington/and INSEAD
Krishna Palepu, Harvard University
Zoe-Vonna Palmrose, University of Southern California
Ken. V. Peasnell, University of Lancaster
Mark C. Penno, University of Chicago
Hasan Pirkul, The Ohio State University
Gordon D. Richardson, University of Waterloo
Wayne H. Shaw, Cornell University
Michael D. Shields, San Diego State University
Rajendra P. Srivastava, University of Kansas
Mary S. Stone, University of Alabama
Charles W. Swenson, University of Southern California
Senyo Tse, University of Texas at Austin
William S. Waller, University of Arizona
Gregory Waymire, Emory University
John J. Wild, University of Wisconsin at Madison
Earl R. Wilson, University of Missouri at Columbia
G. Peter Wilson, Harvard University
William F. Wright, University of California at Irvine/Duke University

THE ACCOUNTING REVIEW

A QUARTERLY JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

Editor
A. Rashad Abdel-khalik

Associate Editors

Robert H. Ashton	John S. Hughes
Andrew D. Bailey, Jr.	Richard A. Lambert
Bala V. Balachandran	James C. McKeown
Victor L. Bernard	Ira Solomon
Ronald W. Hilton	Shyam Sunder

VOL. 66 APRIL 1991 No. 2

A FORUM ON COGNITIVE RESEARCH IN AUDITNG

ALISON HUBBARD ASHTON
Experience and Error Frequency Knowledge as Potential Determinants of Audit Expertise 218
DAVID M. FREDERICK
Auditors' Representation and Retrieval of Internal Control Knowledge 240
ERIC E. SPIRES
Auditors' Evaluation of Test-of-Control Strength 259
ROBIN M. HOGARTH
A Perspective on Cognitive Research in Auditing 277

WILLIAM M. CREADY and PATRICIA G. MYNATT
The Information Content of Annual Reports: A Price and Trading Response Analysis 291
KRISHNAGOPAL MENON and DAVID D. WILLIAMS
Auditor Credibility and Initial Public Offerings 313
JOHN WURST, JOHN NETER, and JAMES GODFREY
Effectiveness of Rectification in Audit Sampling 333

SHORTER ARTICLES

ALAN ROSHWALB and ROGER L. WRIGHT
Using Information in Addition to Book Value in Sample Designs for Inventory Cost Estimation 348
RICHARD K. FLEISCHMAN and LEE D. PARKER
British Entrepreneurs and Pre-Industrial Revolution Evidence of Cost Management 361
DALE MORSE, JENS STEPHAN, and EARL K. STICE
Earnings Announcements and the Convergence (or Divergence) of Beliefs .. 376
BIPIN B. AJINKYA, ROWLAND K. ATIASE, and MICHAEL J. GIFT
Volume of Trading and the Dispersion in Financial Analysts' Earnings Forecasts 389
SCOTT E. STICKEL
Common Stock Returns Surrounding Earnings Forecast Revisions: More Puzzling Evidence 402

Contents continued on Overleaf

BOOK REVIEWS, Shane Moriarity, Editor

Archer, McLeay, and DuFour, *Audit Reports on the Financial Statements of European Multinational Companies: A Comparative Study* .... MOHAMED E. A. HUSSEIN 417

Bannock, *Taxation in the European Community: The Small Business Perspective* .... MARK A. SEGAL 418

Berg, Daley, Gigler, and Kanodia, *The Value of Communication in Agency Contracts: Theory and Experimental Evidence* .... RUSSELL LUNDHOLM and NAHUM MELUMAD 419

Canadian Institute of Chartered Accountants, *Accounting and Reporting for Physical Assets by Governments* .... JESSE W. HUGHES 420

Chant, *The Financial Statement Presentation of Corporate Financing Activities* .... ROSANNE M. MOHR 421

Cowell, *Cheating the Government: The Economics of Tax Evasion* .... ROBERT J. ROLFE 422

Edwards and Connell, *Expert Systems in Accounting* .... JAMES V. HANSEN 423

Needles and Zimmerman, Editors *Comparative International Accounting Educational Standards* .... THOMAS B. CLEVENGER 424

Nobes, *Interpreting European Financial Statements: Towards 1992* .... LEE H. RADEBAUGH 425

Winsten, *Accounting Software and the Microcomputer: A Practical Guide to Evaluation and Implementation* .... PHILLIP D. HARSHA 426

List of Books Received 427

Information for Authors 429

PAST EDITORS

1926–1929	William A. Paton
1929–1943	Eric L. Kohler
1944–1947	A. C. Littleton
1948–1949	Robert L. Dixon
1950–1959	Frank P. Smith
1960–1962	Robert K. Mautz
1963–1964	Lawrence L. Vance
1965–1967	Wendell Trumbull
1968–1970	Charles H. Griffin
1971–1972	Eldon S. Hendriksen
1973–1975	Thomas F. Keller
1976–1978	Don T. DeCoster
1978–1983	Stephen A. Zeff
1983–1987	Gary L. Sundem
1987–1989	William R. Kinney, Jr.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 1–21

The Valuation of R&D Firms with R&D Limited Partnerships

Terry Shevlin
University of Washington

SYNOPSIS: This paper investigates whether capital market investors, in assessing market values of R&D firms' equity, view R&D limited partnerships (LPs) as increasing both the assets and liabilities of the R&D firms. The contract terms between the R&D firm and the LP suggest interpreting the LP as a call option held by the R&D firm and using option pricing theory to estimate the asset (the present value of the LP-funded R&D project) and liability (the present value of the exercise price) components of the option. Estimates of the LP variables are derived from information provided in footnote disclosures by R&D firms. However, several important assumptions are necessary to derive these estimates because the footnote disclosures do not provide all the data necessary to calculate the option values.

The estimates of the LP variables are included as explanatory variables in a cross-sectional market value regression model that is based on the balance sheet identity. Other variables included in the regression model are the assets, liabilities, and in-house expenditures on R&D reported by the R&D firms in their financial statements. The results are consistent with market participants capitalizing in-house R&D expenditures (though SFAS No. 2 (1974) requires R&D expenditures to be expensed as incurred), viewing the call option feature of the LP as relevant information in assessing the market value of the R&D firm, and utilizing the footnote disclosures to assess the option values.

The results support the SEC and the FASB positions that require R&D firms with LPs to disclose to investors the existence of and the amount raised through the LP, but the results do not necessarily support the SEC and FASB positions that this disclosure should be made on the balance

Editor's Note: This article was honored as the 1990 *Competitive Manuscript Contest Award* winner.

The author would like to thank Gregg Brauer, Dave Burgstahler, George Foster, Thomas Hemmer, Marilyn Johnson, Lauren Kelly, Eric Noreen, Gim Seow, D. Shores, Dan Siegel, and two anonymous referees for helpful discussions and/or comments on earlier drafts of this paper. Sun Ho Park assisted with data collection. Financial support from the KPMG Peat Marwick Foundation and the Accounting Development Fund at the University of Washington is gratefully acknowledged.

Submitted September 1987.
Revisions received July 1988, July 1989, and January 1990.
Accepted May 1990.

sheet. The problems I encountered in estimating the LP variables suggest that more detailed disclosures of the terms of the LP option would help investors to assess the market value of R&D firms. More generally, the results suggest that the FASB should pay particular attention to the types of required disclosures and not be so concerned with whether the disclosures should be on or off-balance sheet in their ongoing consideration of financial instruments and transactions.

Key Words: *Off-balance sheet financing, Option models, R&D limited partnerships, Valuation.*

Data Availability: *Data are available on request from the author.*

BOTH the Securities and Exchange Commission (SEC) and the Financial Accounting Standards Board (FASB) have become increasingly concerned with the growing use of arrangements that create potential commitments that are not reported as liabilities in the primary financial statements. One of the earliest off-balance sheet financing arrangements was the sale and leaseback of real property. Other examples include wholly or partly owned but unconsolidated subsidiaries, factoring of accounts receivables, take-or-pay contracts, and product financing arrangements. More recent innovations are research and development limited partnerships (R&D LPs), which are used as funding vehicles by firms.

The SEC argues that the funds raised through an R&D LP should be reported as venture capital borrowings (debt) on the balance sheet of the R&D firm (see Bettner 1981). The FASB, in SFAS No. 68 (1982), supports the SEC's position but allows R&D firms to report the existence of the LP in the notes to the accounts if the firm is not obligated to repay the limited partners. More than 90 percent of the sample of R&D firms with LPs and with available data disclose the LP in the footnotes to the financial statements rather than as debt on the face of the balance sheet.[1]

In an efficient capital market, relevant information about firms will be reflected in security prices regardless of whether the information is reported on the balance sheet or in the footnotes (see Dyckman and Morse 1986 for further discussion). Under the maintained hypothesis that security markets are efficient, this paper investigates whether investors, in assessing market values of R&D firms' equity, view R&D LPs as assets and liabilities of the R&D firms. The paper interprets the LP as an option held by the R&D firm and uses option pricing theory to estimate its value. The option interpretation is suggested by the contract terms between the R&D firm and the LP. In exchange for the transfer by the R&D firm to the LP of the rights to the base technology, the R&D firm receives the right or option to exploit first any technology developed by the LP. This option is exercised when the R&D firm pays (either through royalties or the issuing of shares to the limited partners) to exploit the developed technology. The value of the LP call option to the R&D firm may be decomposed into the present value of the underlying R&D project financed by the LP (an asset) less the present value of the payments by the R&D firm to the limited partners if the firm exercises its option (a liability).

---

[1] The use of footnote disclosures is consistent with both tax and off-balance sheet reasons for the R&D firm to use an LP to fund R&D. See Shevlin (1987) for further analysis and empirical evidence consistent with these factors.

Viewing the LP as an option held by the R&D firm facilitates estimating the asset and liability components of the LP call option. These estimates are included as predictor variables in a cross-sectional market value regression model that is based on the balance sheet identity. Other variables included in the regression model are the assets, liabilities, and in-house expenditures on R&D reported by the R&D firms in their financial statements. In-house R&D expenditures are included to test the hypothesis that investors view these expenditures as positive net present value projects that increase the market value of the firm even though SFAS No. 2 (1974) requires the expenditures to be expensed as incurred.

Estimates of the asset and liability components of the LP are derived from information provided in footnote disclosures by R&D firms. Several important assumptions are necessary to derive these estimates because the footnote disclosures do not provide all the data necessary to calculate the option values. Thus the tests are joint tests of whether investors view the LP as a call option with value to R&D firms (via the asset and liability components) as well as tests of the appropriateness of the assumptions used in estimating the LP variables.

The results are consistent with market participants capitalizing in-house R&D expenditures, viewing the call option feature of the LP as relevant information in assessing the market value of the R&D firm, and utilizing the footnote disclosures to assess the option values. The regression coefficients on the LP variables are significantly different from zero consistent with investors viewing R&D LPs as giving rise to both a debt obligation and an asset of R&D firms. These results support the SEC and FASB positions that require R&D firms with LPs to disclose the existence of the LP to investors but do not necessarily support their positions that this disclosure should be on the balance sheet. The magnitude of the LP variables and the regression coefficients are somewhat sensitive to the assumptions underlying their measurement. The problems encountered here in estimating the LP variables could be used to support the argument that more detailed disclosures of the terms of the LP option would help investors to assess the market value of R&D firms. These results suggest that the FASB should pay particular attention to the types of required disclosures (and not be so concerned with whether the disclosures should be on or off-balance sheet) in their ongoing consideration of "how to account for financial instruments and transactions that seek to transfer market and credit risks—for example, futures contracts, interest rate swaps, options, forward commitments, nonrecourse arrangements, and financial guarantees—and for the underlying assets or liabilities to which the risk-transfering items are related." (FASB Status Report No. 200 1989, 4).

The remainder of the paper is organized as follows. The next section briefly outlines the structure of R&D LPs and develops the call option interpretation. Section II presents the cross-sectional valuation methodology, discusses the measurement of variables, and presents some descriptive statistics on the sample. Section III reports the results, and a brief summary and some implications of the results are presented in section IV.

I. The LP as an Option Held by the R&D Firm

The Structure

Before discussing the R&D LP as an option, it is necessary to provide a brief description of the structure of an R&D LP. (For a fuller discussion see Shevlin 1986 and

Spohr and Wat 1983.) An R&D firm (the sponsor) forms a limited partnership with investors (the limited partners). In return for a general partner's interest (i.e., managerial control) and the right to purchase any developed technology, the sponsor transfers the rights to an idea (the base technology) to the partnership. The R&D work is contracted out by the LP to an R&D contractor (often the sponsor or an affiliate of the sponsor). The investors put up cash in return for a limited partner's interest; they enjoy limited liability provided they do not engage in the management of the LP. In return, the limited partners may deduct any partnership losses in their personal tax returns and the partnership owns any successfully developed technology. Income may be earned either through licensing the developed technology to the sponsor (a royalty partnership) or by forming a new firm to exploit the technology. This new firm could be a joint venture between the sponsor and the limited partnership (a joint venture partnership), or it could be via an issue of equity in the sponsor firm to the limited partners (an equity partnership).

The payments by the LP to the R&D contractor, prior to the LP earning income, lead to partnership losses that give rise to tax deductions available to the limited partners under §174 of the Internal Revenue Code. To obtain the deduction, the limited partners must bear the risk that the R&D may not successfully develop any patentable ideas or saleable technology. That is, the limited partners' capital must be at-risk, as defined in §465, for them to take the R&D deduction.[2]

The R&D LP as a Call Option

The position taken by the SEC appears to be that the total dollar amount invested by the limited partners in the LP (the LP amount) should be accounted for as debt by the R&D firm. The FASB in SFAS No. 68 (1982) recognized that the amount of the debt need not equal the amount raised (see, e.g., p. 3) but gave little guidance as to how to estimate the liability amount. Interpreting the R&D LP in an option pricing framework provides a formal structure to estimate the present values of the expected inflows and outflows of the R&D firm arising from the LP. As an option, the LP increases the value of the firm's equity when the option has value. The option has value when the asset underlying the option (the R&D project) exceeds the exercise price (the liability to the R&D firm). The call option feature arises because in exchange for the transfer by the R&D firm to the LP of the rights to the base technology, the R&D firm receives the right or option to exploit first any technology developed by the LP. This option is exercised when the firm agrees to purchase or pay royalties to exploit the developed technology.

More formally, assuming a royalty partnership structure, the option will be exercised by the R&D firm if:

$$\pi R - r^{*} R > 0, \tag{1}$$

---

[2] The Tax Reform Act of 1976 and the Revenue Act of 1978 modified the provision that the partner's basis (the limit on the amount of a partner's share of partnership losses that the partner can take as a tax deduction) could include liabilities in excess of the at-risk contribution to the partnership. Prior to these modifications, limited partners could increase their partnership basis through nonrecourse borrowing (i.e., borrowing for which the investor has no personal liability) giving rise to claims of 200–300 percent tax losses on the limited partner's capital contribution. The Tax Reform Act of 1986 classified gains and losses from limited partnerships as passive and not available to be used to offset active income of the limited partners. This reduced the attractiveness of R&D LPs as a tax shelter to wealthy investors and effectively eliminated the formation of any new R&D LPs.

where π is the expected net profit margin per dollar of revenue; R is the present value of the total revenue expected from commercial exploitation of the technology developed by the LP; and r^x is the royalty rate per dollar of revenues so that $r^x R$ is the present value of the total expected royalty payment to the limited partners. It is important to note that the payout structure details, including the royalty rate r^x, are contractually agreed upon in the LP agreement at the inception of the LP but that the revenues are a stochastic variable at that date. If the payout price was to be negotiated after the R&D stage, there would not be an option because presumably the LP would charge the market price at that date for the developed technology. Note further that the LP agreement also specifies the partnership structure and investors cannot later choose an alternative structure.

Unlike the well-known Black and Scholes (1973) option pricing model, the exercise price that the R&D firm must pay to the limited partners, $r^x R$, is not fixed but rather is a function of the revenues expected to be generated by the R&D project funded by the LP. Thus an option pricing model that incorporates a stochastic exercise price is required. The model developed by Margrabe (1978) is adapted to the R&D LP problem. Margrabe's model solves for the value of an option to exchange one risky asset for another. Examples of such options include exchange offers, standby commitments, and general margin fees.[3]

The general model is:

$$\begin{aligned} w &= x_1 N(d_1) - x_2 N(d_2) \\ d_1 &= [\ell n(x_1/x_2) + \frac{1}{2}\sigma^2 T]/\sigma\sqrt{T} \\ d_2 &= d_1 - \sigma\sqrt{T} \\ \sigma^2 &= \sigma_1^2 - 2\sigma_1\sigma_2\rho_{12} + \sigma_2^2, \end{aligned} \tag{2}$$

where w is the value of the option to exchange x_2 for x_1; x_1 (x_2) is the value of the asset to be acquired (given up); $N(\bullet)$ is the cumulative standard normal distribution function; T is the option term to maturity; σ_1 (σ_2) is the standard deviation of the rate of return on asset 1 (asset 2); and ρ_{12} is the correlation between the returns on the two assets.

Equation (2) may be applied to the LP problem to derive the after-tax value to the R&D firm of the call option feature of the LP (hereafter referred to as the LP call option):

$$\mathrm{LPCO} = (1-\tau)[PN(d_1) - XN(d_2)], \tag{2'}$$

where LPCO is the after-tax value to the R&D firm of the LP call option applying the Margrabe model; τ is the R&D firm's marginal tax rate; P is the value of the R&D project funded by the LP (asset 1); X is the exercise price (asset 2); and the other variables are as defined in equation (2). From equation (1), $P = \pi R$ and $X = r^x R$. An after-tax approach is used to reflect the fact that, as previously discussed, taxes are an important variable in the decision by R&D firms to finance the R&D through an LP. The after-tax version is derived simply as $\max\{0, (1-\tau)[x_1 N(d_1) - x_2 N(d_2)]\} = (1-\tau)\max\{0, w\}$. Note that in this formulation it is assumed for simplicity that the marginal tax rate applicable to the inflows, $x_1N(d_1)$, and outflows, $x_2N(d_2)$, is the same. This assumption is unlikely to hold if the flows occur at different times and the firms' marginal tax rate changes through time.

---

[3] Fischer (1978) develops a closely related model and illustrates its application to the valuation of index bonds relative to bonds denominated in nominal terms. The concepts, model, and variable estimation procedures are similar in spirit to those used by McDonald and Siegel (1986) in valuing the option of waiting to undertake an investment project, and by Paddock et al. (1988) in valuing offshore petroleum leases.

The value of the LP call option to the R&D firm in equation (2) may be decomposed into the after-tax present value of the project or asset ($LPCA \equiv (1-\tau)PN(d_1)$) less the after-tax present value of the expected payout to the limited partners ($LPCD \equiv (1-\tau)XN(d_2)$). This decomposition allows an estimate to be derived of the after-tax values to the firm of the asset and liability components of the LP: $LPCO = [LPCA - LPCD]$.

II. Methodology and Measurement of Variables

The stock market valuation of R&D firms with R&D LPs is examined using a cross-sectional valuation model based on the balance sheet identity (see Landsman 1986).[4]

Market Value Test Methodology

The market value of shareholders' equity, S, may be written as:

$$S_j \equiv A_j^x - D_j^x, \tag{3}$$

where A^x and D^x denote the market value of the assets and liabilities of firm j. This is a market value version of the book value accounting balance sheet identity that shareholders' equity (or net assets) equals total assets less total liabilities.

Both A^x and D^x may be decomposed into the market value of the *reported* assets and liabilities on the balance sheet (denoted A and D, respectively), and *nonreported* assets and liabilities. It is hypothesized that important nonreported assets for R&D firms relate to their R&D activities. R&D expenditures presumably are made in anticipation of future income. Therefore, even though SFAS No. 2 (1974) requires firms to expense their in-house R&D expenditures, their anticipated value should be reflected in investors' assessments of the market value of the firm. Therefore, one nonreported asset for R&D firms could be the market value of in-house R&D expenditures, denoted *IHRD*.[5] The after-tax present value of the R&D funded by the LP, together with the after-tax present value of the expected exercise price to acquire the developed technology, could also be used by investors in assessing the market value of the R&D firm. Including both the *IHRD* and the after-tax LP variables in equation (3) and substituting A and D for A^x and D^x gives:[6]

$$S_j = A_j - D_j + IHRD_j + LPCA_j - LPCD_j. \tag{4}$$

---

[4] An alternative test of the debt equivalence of the debt component of the R&D LP could be constructed based on the methodology presented in Bowman (1980) and Dhaliwal (1986). This approach regresses estimates of firms' systematic risk on accounting betas and financial leverage decomposed into reported debt and nonreported LP debt. This approach was applied here, but the coefficient on the accounting beta was negative (when its predicted value was unity). This may not be too surprising because many of the sample firms are young start-ups for which accounting beta is difficult to estimate reliably, and as this variable is a multiplicative factor in the regression model (see Dhaliwal 1986, 654), this influences all the other variables. The LP debt variable was not significantly different from zero. However, the accounting beta result makes it difficult to interpret the insignificant coefficient on the LP debt variable. Investigation of the applicability of this model to the LP problem is a topic for future research.

[5] Bublitz and Ettredge (1989) provide evidence, via security price revisions, consistent with investors viewing the benefits of R&D expenditures being long-lived.

[6] Some other nonreported assets and liabilities could be omitted from the analysis. For example, two other off-balance sheet financing arrangements previously examined in the literature are leases (Bowman 1980) and

Sample Selection

The sample of firms examined here is a subset of the 103 R&D limited partnerships examined by Shevlin (1986). These 103 observations represent 75 different R&D firms (with some firms entering into more than one LP). Of the 75 firms, 22 are omitted.[7] This results in 53 firms representing 73 LPs: 39 of these firms are traded on the OTC market, 7 on the ASE, and 7 on the NYSE. The names of the sample firms are listed in appendix B.

Table 1 presents data on the time period and industry composition of the sample. Panel A presents the sample firms classified by the year in which the LPs are formed. In the case of firms entering into multiple LPs, the firm is classified in the year in which its first LP is formed. The sample clusters over 1981 to 1984, which has implications for data availability. Variables are estimated at the end of each fiscal year that a firm is included in the sample. Given that the sample average of the life of the option on the R&D LP is 4.7 years (see below), each firm is included for each of the four years after the date of the formation of the LP if data are available. Data availability results in 145 yearly observations that are used in the cross-sectional regressions below. Panel B indicates the sample is drawn largely from the drugs (including biotechnology), electronics, computing, and medical instruments industries. Examination of the ten R&D firms in SIC 739 (Services—R&D laboratories) indicates that many of these firms are not simply service laboratories but are full-fledged companies conducting R&D on their own account (as well as possibly on contract for clients).

Measurement of Non-Option Variables

The market value of equity, S, is estimated as the fiscal year-end share price times the number of common shares outstanding. The book value of reported assets is used as a proxy for the market value of reported assets, *A*. Landsman (1986) also uses a book value (historical cost) proxy for the market value of assets and liabilities. When Landsman substitutes current cost data for the book value of assets, his results are consistent with the current cost data adding (rather than reducing) measurement error relative to historical assets. Many of the R&D sample firms do not disclose current cost data as they are below the firm size cutoff of SFAS No. 33 (1979). For these reasons, book value of assets is used here.

The market value of reported debt, *D*, is estimated as the sum of the book value of current liabilities and the market (or present) value of long-term debt (including leases capitalized on the balance sheet). The market value of current liabilities (accounts payable, etc.) are reasonably well-approximated by their book values because of their short maturity dates. For long-term debt, market price data are not available for many issues

---

unfunded pension liabilities (Dhaliwal 1986; Landsman 1986). For the time period analyzed here (1980–1985), capitalized leases are already included in the reported debt variable. Only 14 of the sample firms have nonzero unfunded pension liabilities recorded on the COMPUSTAT tapes (and then not for all sample years). Therefore, unfunded pension liabilities are not included in the empirical tests and are assumed to be reflected in the regression model error term.

[7] Six firms are omitted because they did not disclose the amount raised through the LP, two firms because they reported the LP funds as debt on the face of the balance sheet, eight firms because the required financial statement data are not available, and, finally, six firms because they are private firms that are not traded on the OTC, AMEX, or NYSE.

Table 1
Sample Firms Classified by Year and Industry

Panel A. R&D Sample Classified by Year in Which the Firm's First LP Is Formed:

Year	N
1980	1
1981	8
1982	15
1983	18
1984	10
1985	1
Total	53

Panel B. R&D Sample Classified into Three-Digit Industry Groups:

Industry Name	*SIC*	N
Printing, publishing	270	1
Drugs, pharmaceutical	283	5
Miscellaneous plastic products	307	1
Pollution control machinery	355	1
Office computing	357	2
Radio—receiving	365	1
Radio—transmitting	366	4
Electronics	367	3
Electronic computing equipment	368	8
X-ray equipment	369	2
Industrial measurement instruments	382	2
Surgical and medical instruments	384	8
Services—data processing	737	5
Services—R&D laboratories	739	10
Total		53

of the sample firms.[8] Consequently, price data are collected for the issues listed in *Moody's Bond Record*, and present value techniques are used to estimate market values of all other long-term debt. To calculate present values, details on the long-term debt outstanding (book value, maturity dates, and coupon rates) are collected from *Moody's OTC* and *Industrial Manuals* (1981–1986). The yields on the All Industries Corporate Bonds: Seasoned Issues (published in the *Federal Reserve Bulletin*) are used to discount the future outflows on the debt.[9]

---

[8] *Moody's Bond Record* includes only issues listed on the New York and American Stock Exchanges. The *National Bond Summary* reports a cumulative list of bond market quotations for each six-month period, but a check of this source found many issues of the sample firms are not listed and that infrequent bond trading results in the listed prices being out-of-date.

[9] As noted by Mulford (1985), "The extensive estimation techniques utilized here have the potential for introducing a significant amount of measurement error to the market-value-based debt-to-equity ratios." When the regressions below are estimated using the book value of reported debt similar results are obtained. This may not be too surprising given that current liabilities comprise nearly two-thirds of reported debt of the sample firms and the high correlation of .99 between the book and market value debt measures.

In the theoretical model, *IHRD* is defined as the market (or capitalized) value of firms' in-house R&D. As part of the procedure for estimating the LP call option, estimates of market capitalization factors (k_t) applied to *IHRD* are derived from a control sample of R&D firms.[10] These factors are used to convert the R&D expenses reported each year by the R&D sample firms into an estimate of the capitalized value of in-house R&D. Formally, $IHRD = k_t \cdot$R&D expense for year t.

Measurement of Option Variables

The after-tax LP call option variables (*LPCO*, *LPCA*, and *LPCD*), are functions of four model parameters T, P, X, and σ plus the firm's marginal tax rate τ. All are firm-specific variables not directly disclosed by firms and therefore must be estimated. The definition, notation, and method of estimation of each of the option variable inputs are summarized in table 2 together with the mean and median of the cross-sectional distribution of each variable. At this point it is important to note that while assumptions are necessary to estimate the model inputs, investors also have to make assumptions in deriving their own estimates. Thus in making assumptions here to derive estimates I am attempting to replicate investors' estimates. There exists measurement error to the extent that investors' estimates are not replicated. That is, the measurement error is not with respect to the true (but unknown) value of the LP as an option, but rather with respect to investors' estimates impounded in security prices. The sensitivity of the results to different assumptions is examined below, and the results are found to be robust across the assumptions examined.

The term to maturity, T, is estimated as the term of the R&D project disclosed in the LP prospectus or financial statement footnotes of the R&D firm plus an estimate of the average interim license period, i.e., the period during which the R&D firm assesses the developed technology before deciding whether to purchase it.[11] A two-year period is used as an estimate of the interim license period because it is the average of the figures disclosed in the few LP prospectuses available to the author.

The value of the LP-funded R&D project, P, is estimated as the market value capitalization factor k_t times the dollar amount expected to be spent by the LP on R&D. This amount is assumed to be equal to the amount invested by the limited partners less ten percent representing payment of organizational fees.[12] Estimation of k_t, using a control sample of R&D firms, is discussed in appendix A. The estimates of k_t are after-tax estimates and are restated to a pretax basis when estimating P. Thus, $P = (k_t/(1-\tau)) \cdot (0.90$ of the LP amount).

The exercise price, X, is the purchase price to be paid by the R&D firm for the developed technology. This often takes the form of a royalty agreement in which the R&D firm pays an agreed upon percentage of revenues from the sales of products arising

---

[10] Note that the capitalization factor, k_t, is simply an estimate of the factor that converts the dollar amount spent on R&D in a given year into the market value estimate of the increment in equity. This factor represents the ratio of the present value of the expected payoffs to the R&D outlay (the investment) and may be thought of as the present value index. Estimation of k_t is discussed in appendix A.

[11] Note that the firm may exercise its option during this interim license period. However, Margrabe (1978, 180–81) shows that the option is worth more "alive than dead" and therefore will not be exercised prior to the maturity date.

[12] Again, the amount of the organizational fees is the average reported in the LP prospectuses. Further, the LPs do not appear to issue debt securities but to rely solely on the limited partners' contributions or pledges.

Table 2

Definitions, Medians, and Means of Variables Used to Estimate LP Call Option Values (Dollar amounts are shown in millions of dollars)

$$LPCO=(1-\tau)[PN(d_1)-XN(d_2)]$$
$$d_1=[\ell n(P/X)+\tfrac{1}{2}\sigma^2 T]/\sigma\sqrt{T}$$
$$d_2=d_1-\sigma\sqrt{T}$$

Variable	*Definition*	*Proxy*[a]	*Median*[b]	*Mean*
LPCO	Value to the R&D firm of its option to exploit any technology developed by the LP		\$ 3.473	\$ 13.814
LPCA	Present value of the underlying R&D project financed by the LP	$PN(d_1)$	\$ 9.380	\$ 36.012
LPCD	Present value of the exercise price of the option	$XN(d_2)$	\$ 5.716	\$ 22.198
N(•)	Cumulative normal distribution function			
$N(d_1)$			0.622	0.552
$N(d_2)$			0.347	0.492
P	Market value of the R&D project	$\frac{k_r}{1-\tau}$•(0.90•LP amt)	\$ 28.322	\$ 76.110
k_r	R&D market capitalization factor	See text	4.061	4.417
LP amt	Dollar amount contributed by investors in the LP	Actual	\$ 4.680	\$ 10.873
X	Exercise price of the option	r^x•R	\$ 23.963	\$ 83.764
r^x	Royalty rate payable to LP investors	Actual	0.096	0.096
R	Estimated revenues from exploiting developed technology	P/π	\$276.132	\$927.415
π	Profit margin on LP project	Earnings/Sales of R&D firm	0.077	0.083
T	Term to maturity of the option	Actual (years)	3.667	3.524
σ^2	Variance: $\sigma_P^2-2\rho_{PX}\sigma_P\sigma_X+\sigma_X^2$		0.014	0.040
σ_P^2	Variance of the rate of return on P	var(ROA) of R&D firm	0.015	0.043
σ_X^2	Variance of the rate of return on X	r^{x2}var(SOA) of R&D firm	0.001	0.003
$_{PX}$	Correlation between rates of return on P and X		0.523	0.408
τ	Marginal tax rate of the R&D firm	Simulation	0.393	0.330

[a] The proxy column summarizes the procedure used to estimate each option variable. "Actual" denotes data collected from disclosures by the R&D firm.

[b] Based on 53 firms representing 73 LPs. The value of the LP to the R&D firm as an option is estimated for each LP for each year the R&D firm is in the sample. Thus, the means and medians in this table are based on a sample size of 196 LP/years. For firms with multiple LPs in a given year, the option variables are summed across the LPs within this given year for use in the cross-sectional market value regressions below. This results in 145 yearly observations. The call option variables are all stated here on a before-tax basis. For these two reasons, the numbers reported in this table are not directly traceable to those reported in table 3.

from the developed technology.[13] Therefore, $X=r^x R$ where r^x is the royalty rate reported in the LP prospectus or footnote disclosures by the R&D firm, and R is the revenue from the sale of products from the LP-developed technology. From equation

---

[13] This is a simplification because sometimes the royalty period and/or amount is capped. Further, as noted in section I, there are three different partnership structures, which may give rise to different payment methods. Forty-five of the 73 LPs are royalty LPs, 3 are equity, 18 are joint venture, and 7 could not be classified (for lack of data). Of the 18 joint venture LPs, 8 disclosed a royalty rate. Thus, at least 73 percent of the LPs effectively meet the assumption of a royalty LP used in estimating the option value. Sample averages for the firms with data are used as estimates of the parameters for the cases for which data are not available.

(1), $P \cong \pi R$. Given the estimate of P discussed above, an estimate of R is given by P/π where π, the expected net profit margin per dollar of revenue, is estimated as the R&D firm's average pretax profit margin. This assumes the profit margin on the LP-developed technology is equivalent to the existing profit margin of the firm.[14]

The final variable to estimate is σ^2, which is defined in equation (2). Substituting P and X for assets 1 and 2, respectively:

$$\sigma^2 = \sigma_P^2 - 2\sigma_P\sigma_X\rho_{PX} + \sigma_X^2, \tag{5}$$

where σ_P^2 is the variance of the rate of return on the LP-funded R&D project; σ_X^2 is the variance of the rate of return on the asset given up as the exercise price; and ρ_{PX} is the correlation in the return series of the two assets. The return series on the R&D project P (on the exercise price X) is proxied by the R&D firm's return on assets (sales to assets) using annual data for up to 20 years prior to the LP formation date. This assumes each return series approximates that on the LP project. The variances and correlations are then estimated from these return series (with σ_X^2 estimated as r^{*2} times the variance of the return on sales to reflect the fact that the limited partners receive only a portion of the revenues).[15]

Each firm's marginal tax rate, τ, is derived using the simulation procedures outlined in Shevlin (1987). Briefly, the approach involves estimating the marginal tax rate for each firm by estimatng, via simulation procedures, the change in the present value of taxes paid to the taxing authorities from earning an extra dollar of taxable income in the current period.

There are several other points to note with respect to the LP call option. First, the use of the OPM ignores any incentive problems between the R&D firm and the limited partners (see Wolfson 1985; Shevlin 1986). Ignoring incentive problems may result in the OPM overstating the value of the LP call option. These problems are incorporated into the option estimation procedures by examining the sensitivity of the results to lower R&D market capitalization factors (with the results reported below). Second, for firms with multiple LPs, an after-tax LP call option value is estimated for each LP option outstanding at the end of each sample year. These values are then summed to derive an after-tax estimate of *LPCO*, *LPCA*, and *LPCD* at the end of each sample year for each firm. Finally, for firms with missing values of any of the option model parameters the sample averages are used (with the average term to maturity being 4.667 years, the average royalty rate 9.599 percent, and the sales/earnings ratio 12.597).

---

[14] This assumption is difficult to validate (or invalidate) empirically. Ideally data on the payoffs to both in-house and LP-funded R&D are required; however, neither is readily available for individual firms. Few firms disclose in their annual reports, or in other publicly available documents, the payoffs to the limited partners. Note that, if LPs are largely tax motivated as the evidence in Shevlin (1987) suggests, the assumption of equal profitability on in-house and LP-funded R&D may not be unreasonable. However, to the extent low-quality projects are farmed out to LP funding, the profit margin might be expected to be lower on LP-funded R&D (sales to earnings higher) and, hence, the exercise price to be understated using an equal profitability assumption. Decreasing the profit margin by 50 percent to reflect possible differences in profitability between in-house and LP-funded R&D did not change inferences. Further note that for firms with negative earnings, which are mostly young start-ups, a positive ratio is required and the sample average sales/earnings ratio of 12.597 is used.

[15] Doubling the estimated variance to reflect the possibility that more risky projects are funded by LPs results in slightly increased option values, as predicted by the OPM. However, the regression coefficients are virtually identical to those reported below in equation (7).

III. Predictions and Test Results

Multiple regression analysis is used to test the theoretical model in equation (4). The empirical analogue is:

$$S_j = \alpha_0 + \alpha_1 A_j + \alpha_2 D_j + \alpha_3 IHRD_j + \alpha_4 LPCA_j + \alpha_5 LPCD_j + u_j. \tag{6}$$

If the theoretical model is correct and the variables are measured without error, α_0 should be zero, and the regression coefficients should be as follows: α_1 and α_3 equal to 1, and α_2 equal to -1. If investors view the LP as giving rise to an asset and liability of the R&D firm and value these LP variables in an option pricing framework, then the regression coefficients on the LP variables should be nonzero. Further, if the LP variables are measured without error, α_4 (α_5) is predicted to be equal to 1 (-1).

Three sets of tests on the LP variable regression coefficients are conducted. First, if reported assets and liabilities and the LP variables are measured without error and there are no correlated omitted variables, then an F-test should not reject the null hypothesis that $\alpha_1 = \alpha_4$ and $\alpha_2 = \alpha_5$. Second, as book value is used for reported assets and many assumptions are required in deriving market value estimates of reported debt that could result in the coefficients on these variables being different from their respective predicted values of one and negative one, the regression coefficient on the LPCA (*LPCD*) variable is compared to one (negative one). Third, several assumptions are necessary in deriving estimates of the LP variables that may induce measurement error biasing the regression coefficients toward zero (i.e., away from unity). Thus the weaker prediction that $\alpha_4 > 0$ and $\alpha_5 < 0$ is also tested. Finally, any omitted variables are assumed to be uncorrelated with the measurement error and with the included variables and to be reflected in the error term.

A common econometric problem with valuation regression models such as in equation (6) is heteroscedastic disturbance terms. A common technique is to deflate the variables in the model by some measure of firm size under the assumption that the heteroscedasticity is (approximately) linearly related to firm size. The deflator used here is the book value of shareholders' equity, BVE.[16]

The variables in equation (6) are estimated at the end of each of the four fiscal year-ends after the date of the formation of the LP. The 53 sample firms result in 145 yearly observations (as previously noted, some firms have fewer than four years of data available). Panel A of table 3 presents descriptive statistics on the cross-sectional distributions of the BVE deflated variables used in estimating equation (6), as well as the mean

---

[16] There are several points to note with respect to deflation. First, when the regression model is estimated in undeflated form, the slope coefficients are qualitatively similar to those reported in equation (7) below. Second, both the Breusch-Pagan and Glesjer tests (see Johnston 1984, 300–01) reject the null hypothesis of homoscedasticity at the .10 level. Third, deflation by BVE appears to solve the problem of heteroscedastic disturbances. A regression of the absolute value of the residuals from equation (7) below on BVE fails to produce a significant t-statistic on BVE and the R-square is zero. Fourth, when equation (6) is estimated deflating all variables by sales, all regression coefficients increase in absolute terms, but the statistical significance and inferences are unchanged. Fifth, Landsman (1986), instead of simply deflating by sales, generalizes a technique by Park (1966) to deflate the variables. A similar technique, using BVE, produces qualitatively similar results to those reported below. Therefore, only the results for the simpler deflation technique are reported here. Finally, Landsman's (1986) concern that there would be almost near perfect collinearity between A_j, D_j, and the intercept term in the regression if BVE is used as a deflator (see Landsman 1986, 672, n. 19) suggests that he does not deflate the intercept term. However, in weighted least squares, the intercept term should also be deflated (see Johnston 1984, 302).

Table 3
Descriptive Statistics on the Variables Used in the Empirical Tests of Equation (6)[a]

Panel A. Cross-Sectional Distributional Statistics:

	Deflated by BVE					*Undeflated ($)*[b]	
Variable	*.10*	*.50*	*.90*	*Mean*	*Std. Devn.*	*Mean*	*Std. Devn.*
S	1.09	3.07	6.54	4.10	5.53	214.54	413.83
A	1.17	1.64	3.20	2.22	2.38	291.90	1,226.88
D	0.12	0.53	1.89	1.00	1.50	167.93	857.91
IHRD	0.14	0.58	2.27	1.11	1.98	72.53	205.61
LPCO	0.00	0.13	0.74	0.79	5.42	11.10	25.81
LPCA	0.00	0.40	1.81	1.35	6.47	28.53	51.81
LPCD	0.00	0.25	1.11	0.56	1.23	17.43	37.55
P	0.20	0.96	4.40	2.46	6.90	60.60	98.18
X	0.13	0.78	6.24	2.47	4.40	66.45	125.15
LP amt	0.03	0.16	0.94	0.45	0.93	9.58	13.98

Panel B. Product Moment Correlation Matrix:[c]

Variables	S	*A*	*D*	*IHRD*	*LPCO*	*LPCA*	*LPCD*
S	1.000						
A	0.491	1.000					
D	0.498	0.934	1.000				
IHRD	0.739	0.506	0.670	1.000			
LPCO	0.877	0.341	0.400	0.748	1.000		
LPCA	0.873	0.374	0.419	0.743	0.994	1.000	
LPCD	0.725	0.465	0.446	0.612	0.823	0.879	1.000

[a] Distribution statistics based on 145 yearly observations for 53 sample firms. All variables are deflated by the book value of equity BVE except for the last two columns. S = value of stockholder's equity; A = book value of total assets; D = market value of debt; $IHRD = k_t \cdot$ in-house R&D expenditure, where k_t is an estimated R&D market capitalization factor in period t; $LPCO$ = after-tax value of limited partnership to the R&D firm estimated using Margrabe (1978) call option pricing model (equation (2) in text); $LPCO = LPCA - LPCD$; $LPCA$ = after-tax present value of the LP R&D project; $LPCD$ = after-tax present value of the exercise price of the call option; $P(X)$ = value of LP R&D project (exercise price) before applying option valuation formula; and LP amt = dollar amount raised from investors in the LP.

[b] The mean and standard deviation of the undeflated variables are in millions of dollars.

[c] All variables are deflated by the book value of equity. Cut-offs for five percent significance levels for correlation coefficients for a sample size of 145 are: for a two-tailed test $|r| > 0.163$ and for a one-tailed test $|r| > 0.137$. NC: not calculated.

and standard deviation of the undeflated variables. Note that in 71 of the 145 yearly observations, $P < X$. However, when the option characteristics of the LP are incorporated, in only 21 cases is *LPCO* effectively zero.[17] This reduction in "out-of-the-money" options reflects the positive influence on option values of both the time to maturity and the variance of returns. Finally, the magnitudes of the bivariate correlations suggest the variables (not surprisingly) are collinear, which affects the efficiency of the regression coefficient estimates.

---

[17] The zero value arises because the option is so far "out-of-the-money" that *LPCA* only exceeds *LPCD* at the fourth or fifth decimal point.

The coefficients (t-statistics) when equation (6) is estimated on the data set of 145 yearly observations are:

$$\begin{array}{lllllll}
S_j = & 7.050 & +2.544A_j & -3.887D_j & +1.058IHRD_j & +0.840LPCA_j & -1.884LPCD_j \\
t(\alpha=0) & (3.896) & (11.886) & (-9.366) & (5.365) & (10.492) & (-5.185) \\
t(\alpha=1) & & (7.214) & (-6.956) & (0.296) & (-2.004) & (-2.433)
\end{array} \qquad (7)$$

with an adjusted R square of 0.837.[18] The intercept term is significantly greater than zero, which is consistent with an excess of nonreported assets over nonreported liabilities that are independent of the reported assets and liabilities. The nonzero intercept suggests the empirical model may be misspecified, and therefore the results should be viewed with caution. The regression coefficients on the reported assets and liability variables are significantly different from zero in their predicted directions (as indicated by the first row of t-statistics). However, as reported in the second row of t-statistics, the coefficients are also significantly greater, in absolute terms, than unity, which is consistent with the estimates of these variables systematically understating their market values and/or the systematic existence of other (than R&D) nonreported assets and liabilities. Other nonreported assets might include the human capital of the R&D firms (although this might be also partially reflected in the regression coefficient on the *IHRD* variable). The coefficient on the *IHRD* variable is significantly greater than zero consistent with the market capitalizing firms' R&D expenditures. Furthermore, it is not significantly different from its predicted value of unity, which is consistent with the market capitalization factors on *IHRD* not being significantly different between the control and R&D LP samples.

The regression coefficients on the LP variables are significantly different from zero in their predicted directions. However, both coefficients are significantly different, in absolute terms, from unity. The F-test rejects the null hypothesis that the regression coefficients on the LP asset and liability variables are equal to the coefficients on the reported asset and liability variables, respectively. (The F-statistic equals 34.096 with $p<.01$.) This rejection appears due to the regression coefficients on the reported asset and liability variables being significantly greater, in absolute terms, than their predicted values of unity rather than by the LP variables. The significant nonzero regression coefficients provide reasonably strong evidence consistent with capital market investors viewing LPs as a call option when assessing the market value of the equity of R&D firms. Further, the LP option leads to an increase in the market value of equity, on average, of $8.026 million or 3.7 percent of equity.[19]

Sensitivity Analysis

The above results suggest that the empirical model is misspecified, since the intercept term is nonzero and the regression coefficients on the reported asset and liability variables are significantly greater, in absolute terms, than unity. This section provides some further analysis of the regression model in an attempt to improve the specifica-

---

[18] Note the much larger adjusted R-square for the LP sample compared to the control sample R-square (reported in appendix A) of less than .10. Both trimming and deleting observations (at plus 3 standard deviations) in the control sample marginally increased the R-squares; in the LP sample, the R-squares were reduced to approximately .57 by these procedures. Landsman (1986) reports adjusted R-squares around 0.40.

[19] This estimate is derived as $(\hat{\alpha}_4 \cdot \overline{LPCA} + \hat{\alpha}_5 \cdot \overline{LPCD}) \cdot \overline{BVE} = (0.840 \cdot 1.350 - 1.884 \cdot .559) \cdot 101.644$.

Table 4

Cross-Sectional Regressions of Market Value of Equity on Total Reported Assets, Total Reported Debt, In-House R&D Expenditures, Estimated Value of the LP-Funded R&D, and the Estimated Value of the Call Option Exercise Price

$$S_j=\alpha_0+\alpha_1A_j+\alpha_2D_j+\alpha_3IHRD_j+\alpha_4LPCA_j+\alpha_5LPCD_j+u_j$$

Predicted Sign	α_0	α_1 +	α_2 −	α_3 +	α_4 +	α_5 −	Adj. R^2	F-Stat.[a]
1. Margrabe[b]	7.050	2.544	−3.887	1.058	0.840	−1.884	0.837	34.096
t-stat. ($H_0:\alpha=0$)[c]	3.896	11.886	−9.366	5.365	10.492	−5.185		($p=.000$)
t-stat. ($H_0:\alpha=1$)		7.214	−6.956	0.296	−2.004	−2.433		
2. Variables trimmed	6.466	3.458	−4.888	0.423	1.238	−2.929	0.573	37.890
t-stat. ($H_0:\alpha=0$)	4.114	13.570	−11.925	2.009	9.441	−7.335		($p=.000$)
t-stat. ($H_0:\alpha=1$)		9.646	−9.485	−2.741	1.816	−3.015		
3. Industry *k*	6.557	2.672	−3.916	1.049	2.455	−5.543	0.859	6.454
t-stat. ($H_0:\alpha=0$)	3.951	13.094	−10.043	5.807	13.356	−7.966		($p=.001$)
t-stat. ($H_0:\alpha=1$)		8.194	−7.478	0.269	7.915	−6.529		
4. k_t•.50	7.050	2.544	−3.887	1.058	1.679	−3.769	0.837	9.747
t-stat. ($H_0:\alpha=0$)	3.896	11.886	−9.366	5.365	10.493	−5.185		($p=.000$)
t-stat. ($H_0:\alpha=1$)		7.215	−6.957	0.295	4.245	−3.809		

[a] The F-statistic tests the null hypothesis that $\alpha_1=\alpha_4$ and $\alpha_2=\alpha_5$.

[b] Estimation based on 53 sample firms with 145 yearly observations. All variables are deflated by the book value of equity. See notes to table 3 for definition of variables. Regression 1 uses the Margrabe (1978) option pricing model to estimate the LP variables LPCA and LPCD. Regression 2 trims the variables at three standard deviations. Regression 3 uses industry estimates of the R&D capitalization factors *k* to estimate the LP variables. Regression 4 uses the time-varying R&D capitalization factors reduced by 50 percent to estimate the LP variables.

[c] The first row of t-statistics tests the null hypothesis that the $\alpha_i=0$ for $i=0$ to 5. The second row of t-statistics tests the null hypothesis that $\alpha_i=1$ for $i=1$, 3, and 4, $\alpha_i=-1$ for $i=2$ and 5.

tion. In addition, the robustness of the results on the LP variables is probed by examining alternative estimates of the LP variables.

Inspection of table 3 suggests the distributions of the variables are all positively skewed (the mean is to the right of the median). To examine the impact on the regressions of extreme observations, the upper-tail of each variable is trimmed at plus three standard deviations from the mean. The results are reported as regression (2) in table 4 (the results using the Margrabe model reported in equation (7) are reproduced as regression (1) in table 4 for ease of comparison).[20] The intercept term continues to be significantly greater than zero, and the regression coefficients on the reported asset and liability variables both increase in absolute terms. Further, the coefficient on the *IHRD* variable is reduced in magnitude and is significantly less than unity. The regression coefficients on the LP variables both increase in absolute terms and are significantly different from unity. Overall, trimming the variables reduces the explanatory power of the model but does not change inferences about the LP variables.

---

[20] Given the dependent variable is also trimmed at plus three standard deviations, the model should be estimated using a truncated regression model (e.g., Tobit). Since only two observations are actually trimmed, for simplicity, ordinary least square estimates are reported in table 4. When the model is estimated using Tobit, the results are qualitatively similar as those using ordinary least squares.

Table 4 also reports the effects of varying the estimates of the R&D market capitalization factors. Recall that the results reported in equation (7) (and regression (1) in table 4) use the estimates of the capitalization factors reported in equation (A2) for a control sample of R&D firms without LPs. The factors are allowed to vary across time but are assumed to be constant across different firms within the sample and across the control and LP samples. Sensitivity to these assumptions is examined by using estimated industry capitalization factors and by reducing the capitalization factors by 50 percent.[21]

To estimate industry capitalization factors, the control sample of firms is partitioned by two-digit SIC and industry variables generated by using dummy variables (for the resulting industry partitions) times the in-house R&D expenditures. The regression coefficients on these industry variables are the estimates of the industry R&D capitalization factors. Note that in this regression, because of sample sizes, the within-industry factors are restricted to be constant across time. An F-test that capitalization factors are constant across industries could not reject the null hypothesis at the five percent level ($F=2.030$, $p=.0594$). Nevertheless, because the significance level is close to five percent, LP option values are estimated using the industry capitalization factors. As reported in regression (3) in table 4, the regression coefficients on the LP variables increase in absolute magnitude and are significantly greater, in absolute terms, than unity. The coefficients continue to be significantly different from the regression coefficients on the reported asset and liability variables. Overall, use of the industry estimates of k appear to result in poorer estimates of the LP variables, possibly because of the suppression of the time series variation in the R&D capitalization factors.

Finally, to reflect both the possibility that lower quality projects (i.e., expected returns less than commensurate with the expected risk) may be farmed out to LPs by R&D firms, and the possibility that incentive problems may arise from using an LP compared to in-house funding of R&D, the time-varying R&D capitalization factors reported in equation (A2) are reduced by 50 percent (a substantial reduction). This results in halving the value of the call option variables.[22] This halving of the magnitudes of the LP variables results in doubling the regression coefficients on the LP variables as reported in regression (4) in table 4. The t-statistics comparing the coefficients to zero (unity) are unchanged (increased).[23]

Overall, the results appear robust across several different specifications and are generally consistent with the balance sheet identity, with the call option approach to valuing the LP, and with security prices reflecting the existence of the LP increasing the asset and liabilities of the R&D firms.

---

[21] The LP variables are also estimated using the capitalization factor of 4.655 reported in equation (A1), which constrains the factor to be constant across time and industries. The regression coefficients on both the LP variables increase in absolute value with the coefficient on the LP asset variable not significantly different from unity.

[22] Examination of the option pricing model in equation (2) together with the methods used to estimate P and X indicates that the call option values are linear functions of the market R&D capitalization factors.

[23] Note that while the Margrabe option model is the conceptually correct model for the problem at hand, it is possible that investors might be using the Black and Scholes (1973, BSOPM) model. If this is the case, using the Margrabe model results in measurement error with respect to investors' estimates. When estimates of the LP variables are derived using the BSOPM, the regression coefficients on the LP variables are, in absolute terms, smaller than those reported using the Margrabe model but still significantly different from zero in their predicted directions. Thus use of the BSOPM results in similar inferences.

IV. Summary

In light of the SEC's and FASB's concerns with off-balance sheet financing in general and R&D LPs in particular, this paper investigates whether capital market investors, in assessing equity market values of R&D firms, view LPs as increasing both the assets and liabilities of the R&D firms. To do so, the paper interprets the LP in an option pricing framework. Viewing the LP as an option allows estimates of both the asset and liability components of the R&D LP to be derived. The regression results are consistent with investors capitalizing in-house R&D expenditures, viewing the call option feature of LPs as relevant information in assessing the market value of R&D firms and utilizing footnote disclosures in assessing market values. The regression coefficients on the reported asset and liability variables are significantly greater, in absolute terms, than unity. This is consistent with systematic measurement error in the estimates of the predictor variables arising from the use of book value for reported assets and other omitted assets (e.g., the R&D firm's human capital) and liabilities. The regression coefficients on the LP variables are significantly different from zero in their predicted directions and are generally different from the predicted theoretical value of one (negative one) for the LP asset (liability), depending on the assumptions used to estimate the LP variables. Nevertheless, the nonzero coefficients on the LP variables are consistent with investors viewing R&D LPs as giving rise to both a debt obligation and an asset of R&D firms and increasing the market value of the equity of the R&D firms.

There are several implications arising from the results of this study. Specific to the R&D LP accounting debate, the empirical results are consistent with the argument that the existing footnote disclosures allow investors to make some estimate of the value of the LP to the firm. However, because the estimates of the LP variables and the regression coefficients are relatively sensitive to alternative assumptions about variables that are not consistently disclosed by R&D firms, improvement in these disclosures could be helpful to investors. Such disclosures include the exercise price (details of the sharing rules including royalty rates, estimated lifespan of the option, which consists of the lifespan of the R&D project and interim license periods) and the estimated revenues arising from any developed technology together with profit margins (to allow an estimate of the value of the project). Of course, such disclosures may involve releasing sensitive information to competitors.

More generally, the results indicate that, in addition to the reported assets and debt on the face of the balance sheet, investors use information in the footnotes to help assess the market value of firms. Of course, this finding is not new to the literature. Bowman (1980), Dhaliwal (1986), and Landsman (1986), among others, report similar findings. However, the problems in estimating the value of the LP call option outlined in this paper suggest that the FASB should pay particular attention to the types of required disclosures in their ongoing consideration of financial instruments rather than to whether some financing arrangement should or should not be reported on the balance sheet.

Finally, the results add further support to the empirical usefulness of the balance sheet identity approach, first used by Landsman (1986), to develop a cross-sectional valuation model to address off-balance sheet issues. The significant coefficient on the in-house R&D variable in the market value regressions suggests that investors capitalize R&D expenditures even though firms expense them as required under SFAS No. 2 (1974). This result is consistent with prior results that use different methodologies to examine whether investors capitalize or expense R&D (see, e.g., Dukes 1976, Hirschey and Weygandt 1985, and Bublitz and Ettredge 1989). Future research could employ the regression model used here on an expanded sample of R&D firms over an extended time period to estimate the capitalization rates applied to in-house R&D expenditures by investors. These rates could be estimated allowing for variation across firms or industries and across time, and the variations used to test predictions of economic models of innovation and patent protection. Further, it might be possible under simplifying assumptions to derive estimated useful lives of the R&D expenditures that could be used to determine amortization periods if R&D were to be capitalized.[24]

Appendix A

Estimation of R&D Capitalization Factors

An important variable in estimating the market value of the LP-funded R&D is the capitalization factor, k_t, applied by investors to R&D expenditures. Ideally k_t should be an estimate of the market capitalization factor applied to the LP-funded R&D for each LP. However, LPs are not publicly traded, with the result that there are no market prices and thus these capitalization data are not directly available. An alternative proxy is the capitalization factor on the in-house R&D conducted by the sample R&D firms. However, there are limited yearly data available for the sample R&D firms, resulting in small sample sizes that make it difficult to derive efficient estimates of market R&D capitalization factors.[25] Therefore, a control sample of R&D firms without LPs is used to estimate the market's R&D capitalization factors.

The use of the control sample approach assumes the capitalization factors are constant across the LP and control firms. It further assumes that these capitalization factors are applicable to both in-house and LP-funded R&D. This assumption is not likely to strictly hold as the characteristics (basic versus applied research, lifespan, riskiness, etc.) of the projects may differ between the two funding sources. One reason for these differences, as outlined by Shevlin (1986), is that incentive problems between the R&D firm and the limited partners may influence the types of projects funded through LPs. If "lower quality" projects (where lower quality implies that expected returns are not commensurate with the expected risk) are farmed out to LP funding by R&D firms, then the capitalization factors estimated using *IHRD* will overstate the rates appropriate in the OPM resulting in an overstatement of the value of the LP call option. The effect of overstatement is examined in the results section by reestimating the LP variables using capitalization factors reduced by 50 percent.

An alternative to using k_t estimated from security prices is to attempt to directly estimate the present value of the expected payoffs to the LP-funded R&D. However, this requires knowledge (or assumptions) about the probability, timing, and magnitudes of the payoffs together with an estimate of their riskiness to estimate the discount rate.

---

[24] Using an alternative methodology, Hirschey and Weygandt (1985) and Bublitz and Ettredge (1989) develop some market-based estimates of R&D amortization periods.

[25] A further problem with using the sample R&D firms is that, if the model in equation (6) is theoretically correct, then estimates of *LPCA* and *LPCD* are required in order to obtain an efficient estimate of k_t on the *IHRD* variable. Further, if the LP variables and *IHRD* are correlated, then omitting the LP variables when estimating the capitalization factors results in biased estimates of these factors. The correlation matrix reported in table 3 suggests that this might be a problem as the correlation between *IHRD* and *LPCA* (*LPCD*) is .743 (.812).

Using market estimates of the capitalization factors avoids making arbitrary assumptions. (See Paddock et al. 1988, 482–83 for further comparison of the discounted cash flow approach relative to relying on market prices.)

The control sample of R&D firms without LPs is selected by choosing all firms on the 1987 COMPUSTAT Annual Industrial Tape in the same four-digit SIC codes as the LP sample firms. Excluding the firms with R&D LPs results in 175 firms. These firms are generally larger than the R&D firms in the LP sample, which raises the question of whether the capitalization factors estimated using this sample are appropriate to the LP sample. There are no theoretical reasons why the model in equation (6) should not apply equally to large and small firms in the same industries. As discussed in the paper, this assumption is tested by comparing the coefficient on *IHRD* for the LP sample firms to unity. Generally, the coefficient is not significantly different from unity.

Estimation of equation (6) (without the LP variables) pooling the observations on the 175 control sample firms across years gives the following results (with *t*-statistics in parentheses):[26]

$$\begin{array}{lllll} S_j = -0.573 & +1.982A_j & -2.226D_j & +4.655IHRD_j & \\ \quad\ (-0.736) & (21.727) & (-17.342) & (7.376) & \end{array} \qquad \text{(A1)}$$

with an adjusted *R*-square of 0.065. This model restricts the R&D capitalization factor, the coefficient on the *IHRD* variable, to be constant across time. The reasonableness of this restriction is examined by estimating equation (A1) using an interactive indicator variable, I, for each year 1980 to 1985 on the *IHRD* variable. The estimated model is:

$$\begin{array}{lllll} S_j = -0.589 & +1.961A_j & -2.201D_j & +7.305IHRD_{80,j} & +3.274IHRD_{81,j} \\ \quad\ (-0.758) & (21.389) & (-17.088) & (6.916) & (3.164) \end{array} \qquad \text{(A2)}$$

$$\begin{array}{llll} +3.946IHRD_{82,j} & +6.308IHRD_{83,j} & +4.061IHRD_{84,j} & +3.018IHRD_{85,j} \\ \quad (3.971) & \quad (6.443) & \quad (4.589) & \quad (2.193) \end{array}$$

with an adjusted *R*-square of 0.079. An *F*-test of equality of the individual year coefficients rejects the null hypothesis at the five percent level.[27] Thus, the time-varying capitalization factors reported in equation (A2) are used to estimate the option variables.[28]

There are several points to note with respect to the above results. First, the theoretical model in equation (6) predicts the coefficients on A_j and D_j to be +1 and −1, respectively. The estimates on both the asset and debt variables are significantly greater, in absolute value, than their predicted values. This may be due in part to systematic measurement error in that book value is used for both reported assets and debt for the control sample instead of market values, and there may exist nonreported assets and liabilities other than the potential R&D assets that are not included in A and D. Second, the coefficients on the $IHRD$ variables are predicted to be greater than unity based on the hypothesis that the capital market capitalizes firms' R&D expenditures (because the value of the average R&D project is predicted to exceed its cost). The significant positive coefficients in equations (A1) and (A2) are consistent with this prediction.

---

[26] As discussed in more detail for the LP sample, each of the variables in equation (A1) is deflated by the book value of equity to control for heteroscedasticity. Similar coefficient estimates to those reported in equations (A1) and (A2) are derived when sales is used as the deflator.

[27] The estimated *F*-statistic is 3.618 with (5,828) degrees of freedom. Note further that some variable distributions display positive skewness; the above results are based on trimming observations greater than plus three standard deviations. Estimation of equation (A2) without trimming increases the variation in k_t across years. For example, the estimated yearly coefficients on the *IHRD* variable for 1980 through 1985 are 8.838, 3.046, 3.638, 5.823, 1.561, and 1.865, respectively. Using these estimates of k_t to estimate the LP variables results in virtually identical regression results to those reported in equation (7).

[28] Note that while the capitalization factors are allowed to vary across time, they are assumed to be cross-sectionally constant. While it is not possible to directly estimate a capitalization factor for each LP for each year because of data limitations, some evidence on the appropriateness of the assumption is provided in the paper by estimating industry capitalization factors for the control sample of firms.

Appendix B
List of Sample Firms

Company Name	*Cusip*	*Year In Sample*	*Company Name*	*Cusip*	*Year In Sample*
Accuray Corp.	4398	1284–1285	Infodata Systems Inc.	456850	1282–1284
Adac Laboratories	5313	0983–0984	Information Displays Inc.	456704	1281–1283
AEL Industries	1030	0282–0284	Irvine Sensors	99	1084–1085
ALZA Corp.	22615	1283–1285	KLA Instruments Corp.	482480	0682–0685
American Software Inc.	29683	0483–0484	Marquest Medical Prods. Inc.	571431	0383–0384
Anderson Jacobson Inc.	34046	0382–0384	Mathematical Applications	576718	0381–0383
Automatix Inc.	53271	1283–1285	Mentor Corp—Minn.	587188	0381–0384
Autotrol Technology Corp.	52754	1283–1285	Molecular Genetics Inc.	608519	1283–1285
Aydin Corp.	54681	1283–1285	Monoclonal Antibodies Inc.	609755	0383–0384
Becton, Dickinson & Co.	75887	0983–0985	National Patent Development	637130	1284–1285
Bolt Beranek & Newman Inc.	97689	0684–0685	Optical Radiation Corp.	683836	0781–0783
California Biotechnology	99	1182–1185	Plasma-Therm Inc.	99	1284–1285
Centocor Inc.	152342	1284–1285	Policy Management Systems	731108	1283–1285
Cetus Corp.	157199	0683–0685	Quixote Corp.	749056	0683–0685
Control Data Corp.	212363	1282–1284	Rexon Inc.	99	0982–0984
Cordura Corp.	218661	1282–1284	Scientific Software	99	1281–1283
Denelcor	99	1283–1284	Scimed Life Systems Inc.	808614	0283–0284
Diasonics Inc.	252836	1283–1284	Storage Technology Corp.	862111	1281–1283
Digilog Inc.	253806	0983–0985	Syntex Corp.	871616	0783–0785
Emerson Radio	291087	0381–0383	Thoratec Laboratories Corp.	885175	1281–1282
Forest Laboratories Inc.	345838	0384	Trilogy Ltd.	896194	1283–1284
Genentech Inc.	368710	1282–1285	Ultimate Corp.	903848	0482–0484
Genetic Systems Corp.	371853	1282–1284	University Patents Inc.	914802	0782–0784
Genisco Technology	372298	0981–0983	Velo Bind Inc.	922575	1282–1284
GTI Corp.	362360	1280–1282	Ventrex Laboratories	99	0982–0984
Hybritech Inc.	448607	1282–1284	Xebec	99	0981–0984
Immuno Nuclear Corp.	452529	0780			

References

Bettner, J. 1981. Changes in accounting for R&D tax shelters could cause some companies to drop deals. *The Wall Street Journal*: 44.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81 (May/June): 637–54.

Bowman, R. G. 1980. The debt equivalence of leases: An empirical investigation. *The Accounting Review* 55 (April): 237–53.

Bublitz, B., and M. Ettredge. 1989. The information in discretionary outlays: Advertising, research and development. *The Accounting Review* 64 (January): 108–24.

Dhaliwal, D. S. 1986. Measurement of financial leverage in the presence of unfunded pension obligations. *The Accounting Review* 61 (October): 651–61.

Dukes, R. E. 1976. An investigation of the effects of expensing research and development costs on security prices. *Proceedings of the Conference on Topical Research in Accounting*, edited by Micheal Schiff and George Sorter. Ross Institute of Accounting Research, New York University.

Dyckman, T. R., and D. Morse. 1986. *Efficient Capital Markets and Accounting: A Critical Analysis.* New Jersey: Prentice-Hall.

The Federal Reserve Bulletin. 1980–87. Various issues.

Financial Accounting Standards Board (FASB). 1974. *Accounting for Research and Development Costs.* Statement of Accounting Standards No. 2 (October): Stamford, CT: FASB.

———. 1979. *Financial Reporting and Changing Prices.* Statement of Accounting Standards No. 33 (September): Stamford, CT: FASB.

———. 1982. *Research and Development Arrangements.* Statement of Accounting Standards No. 68 (October): Stamford, CT: FASB.

———. 1989. *Status Report No. 200.* Financial Accounting Series, No. 080 (April 7): Stamford, CT: FASB.

Fischer, S. 1978. Call option pricing when the exercise price is uncertain, and the valuation of index bonds. *The Journal of Finance* 33 (March): 169–76.

Hirschey, M., and J. J. Weygandt. 1985. Amortization policy for advertising and R&D expenditures. *Journal of Accounting Research* 23 (Spring): 326–35.

Johnston, J. 1984. *Econometric Methods.* New York: McGraw-Hill.

Landsman, W. 1986. An empirical investigation of pension fund property rights. *The Accounting Review* 61 (October): 662–91.

Margrabe, W. 1978. The value of an option to exchange one asset for another. *The Journal of Finance* 33 (March): 177–86.

McDonald, R., and D. Siegel. 1986. The value of waiting to invest. *The Quarterly Journal of Economics* 101 (November): 707–27.

Moody's Bond Record. 1981–86. Moody's Investors Service, Inc.: New York.

Moody's Industrial Manual. 1981–86. Moody's Investors Service, Inc.: New York.

Moody's OTC Industrial Manual. 1981–86. Moody's Investors Service, Inc.: New York.

Mulford, C. W. 1985. The importance of a market value measurement of debt in leverage ratios: Replication and extensions. *Journal of Accounting Research* 23 (Autumn): 897–906.

National Bond Summary. 1981–86. Jersey City: National Quotation Bureau, Incorporated.

Park, R. 1966. Estimation with heteroscedastic error terms. *Econometrica* 34 (October): 888.

Paddock, J. L., D. R. Siegel, and J. L. Smith. 1988. Option valuation of claims on real assets: The case of offshore petroleum leases. *The Quarterly Journal of Economics* 103 (August): 479–508.

Shevlin, T. J. 1986. *Research and development limited partnerships: An empirical analysis of taxes and incentives.* Ph.D. diss., Stanford University.

———. 1987. Taxes and off-balance sheet financing: Research and development limited partnerships. *The Accounting Review* 62 (July): 480–509.

Spohr, A. P., and L. Wat. 1983. *Forming R&D Partnerships: An Entrepreneur's Guidebook.* Deloitte Haskins & Sells.

Wolfson, M. A. 1985. Empirical evidence of incentive problems and their mitigation in oil and gas tax shelter programs. *Principals and Agents: The Structure of Business,* edited by J. Pratt and R. Zeckhauser. Boston: Harvard Business School Press.

657.05
Ae 57 R

P 8120

A FORUM ON
ACCOUNTING DISCLOSURES AND STOCK PRICES

1. Siva Swaminathan
   The Impact of SEC Mandated Segment Data on Price Variability and Divergence of Beliefs
   (Submitted: December 1988. Accepted: September 1990.)

2. Earl K. Stice
   The Market Reaction to 10-K and 10-Q Filings and to Subsequent The Wall Street Journal *Earnings Announcements*
   (Submitted: January 1989. Accepted: September 1990.)

3. Mary E. Barth, William H. Beaver, and Christopher H. Stinson
   Supplemental Data and the Structure of Thrift Share Prices
   (Submitted: November 1988. Accepted: February 1990.)

4. Michael J. Brennan
   A Perspective on Accounting and Stock Prices
   (Commissioned by the Editor. Accepted: September 1990.)

Editor's Note: A "Forum" is a collection of articles around a common theme. Accepted refereed articles are followed by an invited "perspective" authored by experts on that theme. In addition to allowing for synergy and to providing alternative explanations for the results, an objective of the "perspective" is to provide guidance for future research on issues related to that common theme.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 23–41

The Impact of SEC Mandated Segment Data on Price Variability and Divergence of Beliefs

Siva Swaminathan
Michigan State University

SYNOPSIS: This study provides empirical evidence on the economic effects of Securities and Exchange Commission (SEC) mandated segment data. In 1970, the SEC required multisegment firms to report segment revenue and income in their 10-K reports. Previous studies have reported that SEC mandated segment data reduced the systematic risk and improved earnings prediction but had no impact on security prices. However, the security price studies suffered from methodological deficiencies. This study overcomes these deficiencies and reexamines the security price impact of SEC mandated segment data. It also extends the earnings prediction studies by examining the impact of segment data on divergence of beliefs of multiple financial analysts' earnings forecasts, an aspect of earnings prediction not studied by previous research. Disclosure of segment data is conceptualized as providing a more precise information signal about firm value to investors. Using theories developed in the information economics literature (Holthausen and Verrecchia 1990, and others), two hypotheses are derived.

The first hypothesis predicts that price variability around the dates of release of 1970 10-K reports will be higher than price variability around the dates of release of 1969 10-K reports. The second hypothesis predicts that divergence of beliefs in May 1971 (after the release of SEC mandated segment data) will be lower than divergence of beliefs in May 1970 (before the release of such data). Based on theory and evidence in Bhushan (1989), two

Editor's Note: This paper was one of two manuscripts honored by the *Competitive Manuscript Contest Award* for 1989.

This paper is based on my Ph.D. dissertation completed at SUNY-Buffalo. I would like to thank the members of my committee, Lawrence D. Brown (chair), Robert Hagerman, Joseph Weintrop, and David Chen for their help and encouragement. I would also like to thank two anonymous referees and workshop participants at the University of Connecticut for their comments, and Chris Moening for her expert word-processing assistance. I am responsible for all errors.

Submitted December 1988.
Revisions received January and July 1990.
Accepted September 1990.

more hypotheses are derived. The magnitudes of increase in price variability and decrease in divergence of beliefs are hypothesized to be positively correlated with the number of segments. The Patell (1976) methodology is used to measure price variability, and multiple financial analysts' earnings forecasts reported in the Standard & Poor's Earnings Forecaster are used to measure divergence of beliefs. The sample is divided into a control group and an experimental group based on segment data disclosures prior to the SEC mandate.

The empirical evidence shows that, for the experimental group, there is a significant increase in price variability and a significant decrease in divergence of beliefs. There are no such effects for the control group. The evidence also shows that, for the experimental group, the magnitudes of increase in price variability and decrease in divergence of beliefs are directly proportional to the number of segments. There are no such relationships for the control group.

Key Words: *Segment data, Information precision, Price variability, Divergence of beliefs.*

Data Availability: *The author has spent considerable time and personal funds to collect the data used in this study and is working on a related project. The author will make the data available for research purposes after the completion of this project, which should be approximately one year after the publication of this paper.*

IN 1970, the Securities and Exchange Commission (SEC) mandated publicly traded multisegment firms to report segment revenue and income in their 10-K reports (Release No. 34-9000 issued on 21 October 1970). Multisegment firms were required to provide information on segment revenue and income for those segments contributing ten percent or more to consolidated revenues or income. Management of the firms was given discretion as to the definition of segments.[1] This study extends previous research on the economic effects of segment data. These studies can be classified in terms of the (1) impact on earnings prediction, (2) impact on systematic risk, and (3) impact on security prices.[2]

---

[1] Before the 1970 SEC mandate requiring disclosure of segment data in 10-K reports, the SEC had issued Release No. 33-4949 in February 1969 (effective immediately) requiring registrants to disclose segment data on Forms S-1, S-7, and 10 which are filed at irregular intervals with the SEC in connection with the issue of new securities. Furthermore, prior to the 1970 SEC regulation some firms had voluntarily disclosed segment data in their annual and 10-K reports, and some had disclosed segment data in their proxy statements that were filed in connection with mergers or acquisitions. These institutional features have been incorporated into the sample selection and research design of this study. This research focuses on the 1970 SEC regulation because this was the first time that systematic periodic disclosure of segment data had been required by the SEC, and, furthermore, it facilitates the research design and comparison of this study to prior research. In December 1976, the Financial Accounting Standards Board (FASB) issued Statement of Financial Accounting Standards No. 14 (SFAS 14) (effective 15 December 1976), which required companies to disclose segment assets, depreciation and capital expenditures, in addition to segment revenue and income. The economic effects of SFAS 14 are not studied in this paper.

[2] See Mohr (1983) for a detailed literature review of the empirical research on segment reporting.

Collins (1976) studied firms that disclosed segment data for the first time in their 10-K reports of 1970 (as a result of the SEC regulation) and showed that forecasts of future earnings were more accurate for models based on segment revenue and income than for models based on consolidated data. Another approach to predictive ability was taken by Baldwin (1984), who showed that accuracy of the *Value Line* Investment Survey analysts' earnings forecasts improved significantly after the disclosure of SEC mandated segment revenue and income data. Simonds and Collins (1978) reported a significant downward shift in beta for firms that disclosed segment data for the first time in their 1970 10-K reports. Similar results were reported in a follow-up study by Collins and Simonds (1979).[3]

Horwitz and Kolodny (1977) compared an experimental group consisting of firms that disclosed segment data in their 1970 10-K reports and a control group that reported on a consolidated basis in 1970 and reported no differences in the pattern of return residuals between the two groups. Ajinkya (1980) also compared an experimental group that disclosed segment data for the first time in 1970 and a control group that voluntarily disclosed segment data prior to 1970. No differences in risk-adjusted returns were observed between the two groups of firms, in either the pre-SEC period or the post-SEC period. As with Horwitz and Kolodny, Ajinkya interpreted this result as being consistent with the hypothesis that SEC mandated segment data did not have any impact on security prices. Thus, previous research on the economic effects of SEC mandated segment data have reported mixed results; the earnings prediction and systematic risk studies reported significant effects, but the security price studies reported no significant effects.

The purpose of this paper is to examine further the economic impact of the 1970 segment reporting requirement and to attempt to reconcile the results from the various approaches. First, it is likely that Horwitz and Kolodny and Ajinkya did not detect any security price impact because they examined signed risk-adjusted returns. Such a methodology implicitly assumes that segment data had either a systematic positive or negative impact on security prices, but not both. If segment data had positive and negative impacts on the prices of different firms, then these effects are likely to cancel each other across a sample of firms and no significant price effects will be detected. Second, they did not identify the exact date on which the segment data were released to the market. Both studies examined the behavior of the residuals over a long time period before and after the regulation went into effect. In an efficient capital market, the impact, if any, of segment data is likely to be on or around the date of first disclosure of such data.[4] This study attempts to overcome these deficiencies by identifying the exact dates on which 10-K reports containing segment data were released, and by examining price variability, measured by squared return residuals, around those dates. This paper also extends the earnings predictability studies by studying the effect of the 1970 SEC mandated segment data on divergence of beliefs of financial analysts' earnings forecasts.[5]

---

[3] Earlier, Horwitz and Kolodny (1977) had reported no significant change in systematic risk as a result of disclosure of SEC mandated segment data. However, Simonds and Collins identified a number of deficiencies in their sample selection and test procedures. These deficiencies were corrected in the Simonds and Collins study.

[4] The Horwitz and Kolodny study did not include screening of sample firms for voluntary disclosure of segment data prior to the regulation, a deficiency remedied in the Ajinkya study.

[5] Barefield and Comiskey (1975) attempted to study this issue. They found weak evidence that financial analysts had less divergence of beliefs after the SEC mandated segment disclosures. However, this result was based on eight observations and should be interpreted with caution.

The remainder of the paper is organized as follows. Section I develops the hypotheses, section II describes the data, section III discusses the empirical analysis, and section IV presents the summary and conclusions.

I. Hypotheses

If $\tilde{Y}$ is an accounting signal that reveals the value of the firm $\tilde{F}$ with some error (noise) $\tilde{\epsilon}$, then $\tilde{Y}$ can be expressed as:

$$\tilde{Y}=\tilde{F}+\tilde{\epsilon}, \tag{1}$$

with

$$\begin{aligned} E(\tilde{\epsilon})&=0,\\ \mathrm{Var}(\tilde{\epsilon})&=S, \text{ and}\\ \mathrm{Cov}(\tilde{F},\tilde{\epsilon})&=0. \end{aligned}$$

E(•), Var(•), and Cov(•) are the expectation, variance, and covariance operators, respectively. In information economics terminology, S^{-1} is known as the precision of the information signal $\tilde{Y}$.[6]

The SEC argued that different segments of a firm could differ in their profitability, riskiness, and future growth opportunities. Consequently, revenues and income from different segments are likely to have differing implications for future cash flows of the total firm. For example, revenues or income from a high-growth segment are likely to be associated with higher levels of future cash flows compared to revenues or income from a low-growth segment. Furthermore, different risks imply that different discount rates are to be used to discount the cash flows of each segment. In his testimony before the United States Congress, former SEC Chairman Manuel Cohen asserted that segment revenue and income information would enable the investor to analyze more *precisely* the component parts of the total firm (SEC 1966). Stated in terms of equation (1) above, an information signal based solely on consolidated data is likely to have a lower precision (higher S) than would an information signal that includes segment data. Investors can use segment data to analyze each segment separately taking into account differences in risk, profitability, and growth across segments, and to make estimates of firm value more precise, than those based on consolidated data.

Hypotheses 1 and 2

Holthausen and Verrecchia (1990) examined the effects of information precision on price variability and divergence of beliefs.[7] In their model each agent in the market is assumed to receive a public information signal that has a noise term common to all agents and an idiosyncratic noise term unique to each agent. Using a Walrasian market equilibrium, they showed that a decrease in the idiosyncratic noise term (i.e., increase in precision) *increased* price variability and *decreased* divergence of beliefs. The effect of a change in the common noise term on price variability and divergence of beliefs was

---

[6] In some information economics papers, "S" is referred to as the variance of the information signal noise.

[7] The term consensus was used by the authors, which is the inverse of divergence of beliefs. For easier exposition, unless otherwise stated the term *divergence of beliefs* used in the hypotheses section refers to divergence of beliefs of market participants about future cash flows or about firm value. However, it is operationalized as divergence of beliefs of financial analysts' forecasts of earnings. Similarly, the term *price variability* refers to price variability around the date of (and due to) release of an information signal.

ambiguous. However, they argue that the concepts of common and idiosyncratic noise terms need not necessarily be interpreted as public and private information, respectively. When agents receive a public information signal (e.g., SEC mandated segment data), it is likely to have idiosyncratic noise terms because each agent interprets the signal differently due to differences in background, education, ability to process information, etc. Thus, the comparative statics for the idiosyncratic noise term can be used to examine the effects of SEC mandated segment data on price variability and divergence of beliefs.[8]

For price variability, similar comparative static results have been derived in Holthausen and Verrecchia (1988) and Choi and Salamon (1989). In the Holthausen and Verrecchia (1988) model, price of the risky asset is set equal to the conditional expectation of the value of the asset given all (public) information, and it is set by a risk-neutral competitive market maker. They showed that in the single risky asset case, price variability increased unambiguously as the precision of information increased. Similar results were obtained for the multiple assets case by placing certain restrictions on the intertemporal correlation in the information disclosures and on the cross-sectional correlation in asset returns.[9] Choi and Salamon (1989) derived a similar result using the Capital Asset Pricing Model (CAPM) framework. Using a public information, multifirm setting, they showed that the higher the information precision, the larger the price variability when the information is released. However, neither study examined the relationship between precision and divergence of beliefs.

The relationship between information precision and divergence of beliefs was examined by Cukierman and Wachtel (1979) and Cukierman and Givoly (1982). Cukierman and Wachtel used a rational expectations framework to show that an inverse relationship exists between divergence of beliefs and information precision. Cukierman and Givoly used a Bayesian model to show that the larger the precision of information, the lower the divergence of beliefs among agents attempting to forecast the future earnings number. They also show that more precise information reduces uncertainty about the future earnings number, which in turn reduces divergence of beliefs. This relationship between future earnings uncertainty and divergence of beliefs has been documented in the empirical literature (Ajinkya and Gift 1985; Brown et al. 1987; Daley et al. 1988; Elton et al. 1984). If the risk, profitability, and growth characteristics of a firm are different across segments and if investors have access only to consolidated data, then providing segment data to investors is likely to reduce uncertainty about future earnings. It follows from the Cukierman and Givoly model that if uncertainty about future earnings is reduced, divergence of beliefs among investors is reduced.[10]

Based on the above discussion, on the relationship between precision and price variability and between precision and divergence of beliefs, and by conceptualizing segment data as providing more precise information, the following two hypotheses are formulated:

---

[8] Although not formally shown, they state that these comparative static results would hold equally well in a partially revealing rational expectations framework.

[9] In an earlier unpublished version of this paper (Holthausen and Verrecchia 1981), they showed that this result holds unambiguously in a rational expectations framework with a single risky asset.

[10] Cukierman and Wachtel and Cukierman and Givoly did not study the relationship between precision and price variability.

H1: Price variability around the date of release of 10-Ks that contain SEC mandated segment data is *greater* than price variability around the date of release of 10-Ks that do not contain SEC mandated segment data.[11]

H2: The release of SEC mandated segment data in 10-K reports *decreased* divergence of beliefs among market participants.[12]

Hypotheses 3 and 4

It is possible that the increase in information precision as a result of disclosure of segment data is related to the amount of private information collected prior to the public disclosure of SEC mandated segment data. Bhushan (1989) developed a model that examined the factors determining private information collection. He showed that the higher the marginal cost of collecting an additional unit of precision (from private information), the smaller the amount of private information that will be collected. He then argued that the marginal cost of collecting an additional unit of precision is directly proportional to the number of segments in a firm. As a result, Bhushan hypothesized that the greater the number of segments, the smaller the amount of private information collected. The empirical evidence provided by Bhushan is consistent with this hypothesis. Since less private information is collected for firms with more segments, the private information for such firms will be less precise. Consequently, the *increase* in precision as a result of public disclosure of SEC mandated segment data will be greater for firms with more segments. Conditional on the results of hypotheses H1 and H2, the magnitude of increase in price variability and magnitude of decrease in divergence of beliefs will be larger for firms with more segments. This leads to the formulation of hypotheses H3 and H4:

H3: The magnitude of increase in price variability as a result of SEC mandated segment disclosure is *positively correlated* with the number of segments of a firm.

H4: The magnitude of decrease in divergence of beliefs as a result of SEC mandated segment disclosure is *positively correlated* with the number of segments of a firm.

II. Data

Multiple financial analysts' forecasts of earnings are used as a surrogate for divergence of beliefs of market participants.[13] The Standard & Poor's Earnings Forecaster

---

[11] Hypothesis H1 can also be derived by conceptualizing the disclosure of segment information as movement from a coarser to a finer information environment. A finer information environment is defined as a system in which the set of available state descriptors is a superset as compared with some alternative coarser information environment. Segment versus consolidated information is a good example of a coarser versus finer information environment (Ajinkya 1980; Mohr 1983). Using a CAPM framework Ohlson (1979) showed that price variability is higher in a finer information environment, but did not study the relationship between finer information and divergence of beliefs. The concepts of precision and finer information environment can be said to be inextricably linked. A finer information environment is likely to result in a more precise information signal.

[12] Hypothesis H2 can also be derived by conceptualizing the disclosure of segment information as increasing the *quantity* of public information. Barry and Brown (1985) show, using a CAPM based model, that divergence of beliefs is a monotonically decreasing function of the quantity of public information disclosed. They did not study the relationship between quantity of information and price variability.

[13] Holthausen and Verrecchia (1990) suggested trading volume as an empirical proxy for divergence of beliefs. However, due to the difficulty in obtaining trading volume data for the 1969–1970 time period, this approach is not used in this study. Previous research has used multiple analysts' earnings forecasts as a proxy for divergence of beliefs (Ajinkya and Gift 1985; Brown et al. 1987; Daley et al. 1988; Elton et al. 1984).

(EF) is used as the source of analysts' earnings forecasts. EF was a biweekly publication of the Standard and Poor's Corporation, listing annual earnings forecasts for about 1,500 companies that included forecasts made by about 70 analysts (EF is now published as the ZACKS Earnings Forecaster).

The Center for Research in Security Prices (CRSP) daily returns tape, which contains individual security returns, is used to calculate price variability. The preliminary sample consists of firms that met the following three criteria:

1. Have at least three analysts' earnings forecasts available in the relevant issues of EF.[14]
2. Have returns on the CRSP tape on and around the dates of release of the relevant 10-K.[15]
3. Have a 31 December fiscal year-end.[16]

The SEC segment regulation became effective for fiscal years ending on or after 31 December 1970. The 10-K reports of 1970 of firms that met all three criteria listed above were examined to check whether they disclosed any segment data. The firms in the sample were carefully screened, using 10-K and annual reports of 1969 and registration and proxy statements filed prior to 1970, for any public disclosure of segment data before the SEC regulation went into effect. Each firm was assigned to a control group or an experimental group based on the level of segment disclosure prior to the 1970 SEC regulation.[17]

The results of the sample selection are presented in table 1. Panel A shows that 380 firms met all three criteria listed above, but for the reasons listed in the table, a final sample of 160 experimental and 101 control firms remained for study and analysis.[18]

III. Empirical Analysis

Test Method (H1)

By sample construction, the firms in the experimental group made their first public disclosure of a portion or all of the SEC mandated segment data in their 1970 10-K reports. For these firms, the 1970 10-K reports contained more segment data (SEC mandated segment revenue and income data) than did their 1969 10-K reports (either no segment data or only segment revenue data). To test hypothesis H1, the difference between price variability surrounding the release of the experimental group's 1969 and 1970 10-K reports are examined and compared to that of the control group. Price vari-

---

[14] Requiring at least three forecasts is a subjective criterion, and involves a tradeoff between having a reasonable sample size and obtaining a meaningful measure of divergence of beliefs. The forecasts were collected from the 22 May 1970 and 21 May 1971 issues of EF. This is discussed in detail in the empirical analysis section.

[15] Sufficient returns around the 10-K filing dates are necessary to estimate the parameters of the market model. This is discussed in the empirical analysis section.

[16] This constraint is imposed to minimize the number of issues of EF that had to be obtained. Most libraries did not subscribe to EF during the 1970–1971 time period. The relevant issues of EF were obtained from the Princeton University Libraries. The 31 December fiscal year-end requirement also simplifies the research design.

[17] Only the year preceding the SEC regulation is examined for voluntary prior disclosure of segment data. This assumes that firms follow a consistent disclosure policy.

[18] For brevity, the empirical results are not reported separately for experimental group 1 and experimental group 2. However, all the empirical tests were done separately for each experimental group, and the results were the same for both groups.

Table 1
Sample Selection

Panel A. Results of Screening:		
	Firms with at least three analysts' earnings forecasts in the Standard & Poor's Earnings Forecaster, the necessary return data on the CRSP tape, and a December fiscal year-end	380
Less:	Firms whose 10-K reports could not be located or the date stamp was unreadable	(12)
Less:	Firms that did not disclose any segment data in their 1970 10-K reports	(77)
Less:	Firms with conflicting events around the 10-K release dates	(18)
Less:	Firms whose segment classification scheme in their 1970 10-K report was different from that in their voluntary disclosure (if any)	(12)
	Final Sample	261
Panel B. Classification of Final Sample:		
Control Group		
	Firms that voluntarily disclosed segment revenue and income in their 1969 10-K reports	88
	Firms that voluntarily disclosed segment revenue and income in their 1969 annual reports	2
	Firms that disclosed segment revenue and income data in any registration statement filed prior to 1970	7
	Firms that disclosed segment revenue and income data in any proxy statement filed prior to 1970	4
	Total in Control Group	101
Experimental Group		
	Firms that had no segment disclosure of any kind prior to 1970 (Experimental Group 1)	77
	Firms that disclosed segment revenue (only) in their 1969 10-K report (Experimental Group 2)	83
	Total in Experimental Group	160

ability is measured as the square of the residual returns from the market model (Sharpe 1964; Lintner 1965):

$$R_{i\tau}=\alpha_i+\beta_i R_{m\tau}+e_{i\tau}, \tag{2}$$

where:

$R_{i\tau}$ = return on firm i for period τ,
$R_{m\tau}$ = return on the market portfolio for period τ,
$e_{i\tau}$ = error term in the regression,
$i=1,\ldots,N$, firm index, and
$\tau=1,\ldots,T$, time index.

The SEC stamp on the first page of the 10-K report is used to identify the event day, the date a 10-K report was filed with the SEC. The event day and the two days preceding and the two days following the event day are referred to as the event period. The estimation period consists of the 100 trading days that preceded the event period.[19] The CRSP daily returns tape was used to gather individual security returns. The value weighted market index was used as a proxy for returns on the market portfolio. Ordinary least squares regression was performed on realizations of $R_{i\tau}$ and $R_{m\tau}$ (estimation period) to obtain estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$ of α_i and β_i, respectively, in equation (2). The estimated coefficients $\hat{\alpha}_i$ and $\hat{\beta}_i$ were used to form predictions of R_{it} (for event period).[20] The residual return was defined as:

$$u_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt}; \qquad t = -2, -1, 0, +1, +2. \tag{3}$$

Following Patell (1976), price variability during the event period for firm i was calculated as:

$$U_{it} = \left[\frac{u_{it}^2}{C_{it} s_i^2}\right]\left[\frac{T_i - 4}{T_i - 2}\right]; \qquad t = -2, -1, 0, +1, +2, \tag{4}$$

where:

s_i^2 = the variance of the residuals during the estimation period for firm i;
C_{it} = the increase in variance due to prediction outside the estimation period for firm i on day t; and
T_i = the number of days used in the estimation period for firm i.

The average price variability for each group (control, experimental), for each year's 10-K report (1969, 1970) for day t was calculated as:

$$\bar{U}_t = \frac{1}{N}\sum_{i=1}^{N} U_{it}; \qquad t = -2, -1, 0, +1, +2, \tag{5}$$

where N is the sample size for each group. $\bar{U}_t$ close to one implies no significant change in price variability on day t, and $\bar{U}_t$ greater (less) than one indicates increase (decrease) in price variability on day t (for details, see Patell 1976). The difference between the 1970 and 1969 price variability for firm i on day t was calculated as:

$$CHPV_{it} = U_{it}^{1970} - U_{it}^{1969}, \tag{6}$$

where 1969 and 1970 refer to the 10-K reports of those fiscal years. To standardize $CHPV_{it}$ and make them cross-sectionally comparable, the percentage change in price variability for firm i on day t was calculated as follows:

$$PCHPV_{it} = \frac{CHPV_{it}}{U_{it}^{1969}}. \tag{7}$$

---

[19] The time index t is used to denote the event period and the index τ to denote the estimation period. A minimum of 50 observations were used to estimate the coefficients of the market model. Other estimation periods were also used (50 and 150 days before the event day) and the empirical results were similar.

[20] Summary statistics from the market model regressions were consistent with those reported by previous research. For fiscal 1969 (1970) 10-K reports the mean beta, i.e., systematic risk, was 1.252 (1.131) for the control group and 1.020 (0.809) for the experimental group. The reduction in beta for the experimental group is consistent with the evidence provided by Simonds and Collins (1978) and Collins and Simonds (1979). Price variability was also calculated using a model where the mean of the returns from the estimation period is used as the expectation for the event period. The results using the mean return model were similar to those of the market model.

Hypothesis H1 implies testing whether $PCHPV_i$ for the experimental group is significantly higher than $PCHPV_i$ for the control group.[21]

Test Method (H2)

Financial analysts' earnings forecasts for fiscal 1970, reported in the last EF issued in May 1970 (22 May 1970) were collected for all firms in the sample. Similarly, analysts' earnings forecasts for fiscal 1971, reported in the last EF issued in May 1971 (21 May 1971), after the first disclosure of segment data, were also collected. Since the SEC requires 10-K reports to be filed within 90 days of the fiscal year-end, it is assumed that the May 1971 forecasts are conditional on knowledge of SEC mandated segment data. Most 10-K reports were filed between 21 February and 31 March, although a small percentage of the 10-K reports were filed after the SEC 90-day deadline (31 March). However, all 10-K reports in the sample were filed by end of April (see the appendix). Thus, it is reasonable to assume that analysts' forecasts collected from the 22 May 1970 and 21 May 1971 issues of EF reflect the information contained in the 1969 and 1970 10-K reports, respectively.[22]

Let $FC_{i,j}^m$ be the earnings per share forecast for firm i by analyst j collected from the EF issued in period m (May 1970, May 1971). The following measure of divergence of beliefs is calculated:

$$DB_i^m = \frac{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(FC_{i,j}^m - \overline{FC}_i^m)^2}}{\overline{FC}_i^m}, \tag{8}$$

where:

$$\overline{FC}_i^m = \frac{1}{n}\sum_{j=1}^{n} FC_{i,j}^m$$

and n = number of analysts making the forecast; $n \geq 3$.

DB_i^m is the standard deviation of multiple analysts' earnings forecasts divided by the mean of the forecasts.[23] The difference in divergence of beliefs between May 1970 and May 1971 for firm i was calculated as follows:

$$CHDB_i = DB_i^{\text{May 1971}} - DB_i^{\text{May 1970}} \tag{9}$$

---

[21] All the empirical tests were also conducted on $CHPV_i$ and the results were similar to those of $PCHPV_i$. $PCHPV_i$ was also measured using U_{ik}^{1970} in the denominator and the results were similar.

[22] Multiple analysts' earnings forecasts for a firm, collected from the same issue of EF, do not necessarily have the same forecast date. For any particular firm, earnings forecasts are received by EF from financial analysts at different points in time. EF then publishes biweekly the most recent forecast submitted by each analyst. However, EF gives analysts the opportunity to continually revise their forecasts. If analysts have incentive to submit their best forecast to EF, then it is likely that a forecast reported in EF reflects all the information each analyst has received up to the date of issue of EF, even though the forecasts have been submitted to EF on different dates.

[23] In statistical terminology, standard deviation of a variable divided by its mean is known as coefficient of variation. Previous research has used coefficient of variation of analysts' earnings forecasts as a measure of divergence of beliefs (Ajinkya and Gift 1985; Brown et al. 1987; Daley et al. 1988). In the 22 May 1970 (21 May 1971) issue of EF, firms in the sample were followed, on average, by 4.78 (4.61) analysts for the control group, and 4.32 (4.24) analysts for the experimental group.

Table 2
Preliminary Results—Price Variability and Divergence of Beliefs

Panel A. Price Variability Around 1969 and 1970 10-K Report Release Dates:[a]

		1969 10-K	*1970 10-K*	*1969 vs. 1970 10-K*
	Day	$\bar{U}_t$	$\bar{U}_t$	*Wilcoxon Z*[b]
Control Group[c] *(N=101):*	−2	1.313	1.411	1.35
	−1	1.396	1.489	1.17
	0	1.414	1.495	1.09
	+1	1.167	1.192	0.44
	+2	0.949	0.972	0.50
Experimental Group[c] *(N=160):*	−2	1.341	1.751	2.41**
	−1	1.428	1.801	2.19*
	0	1.447	1.787	1.98*
	+1	1.126	1.202	1.09
	+2	0.946	1.012	0.86

Panel B. Divergence of Beliefs Before and After the Release of SEC Mandated Segment Data:[d]

	May 1970	*May 1971*	*May 1970 vs. May 1971*
	DB	*DB*	*Wilcoxon Z*[e]
Control Group (N=101):	0.0484	0.0482	−0.33
Experimental Group (N=160):	0.0508	0.0279	−4.89**

[a] Day 0 is the date stamped by the SEC on the first page of the 10-K report, indicating date of receipt. $\bar{U}_t$ is the average price variability on day *t*.

[b] Wilcoxon Z is the test statistic from the Wilcoxon Signed Ranks Test, which tests whether U_t for 1969 10-K is equal to U_t for 1970 10-K. **(*) indicates significance at the one percent (five percent) level, one-tailed.

[c] Control group and experimental group are described in table 1.

[d] DB is the average divergence of beliefs of multiple analysts' earnings forecasts. May 1970 and May 1971 indicate dates of issue of the Standard & Poor's Earnings Forecaster from which the forecasts were collected.

[e] Wilcoxon Z is the test statistic from the Wilcoxon Signed Ranks Test, which tests whether DB for May 1970 is equal to DB for May 1971. ** indicates significance at the one percent level (one-tailed).

To standardize $CHDB_i$, the percentage change in divergence of beliefs was calculated as follows:

$$PCHDB_i = \frac{CHDB_i}{DB_i^{May\ 1970}}. \tag{10}$$

Hypothesis H2 implies testing whether *PCHDB* for the experimental group is significantly lower than *PCHDB* for the control group.[24]

Results (H1 and H2)

Table 2 presents preliminary results for hypotheses H1 and H2. The top half of table 2 presents $\bar{U}_t$ (average price variability on day *t*) for the control group and experimental group, for fiscal 1969 and 1970 10-K reports. Table 2 shows that all $\bar{U}_t$ except for day +2 are considerably greater than one for both groups, for fiscal 1969 and 1970. Thus, almost all the capital market reaction to 10-K reports occurs on days −2, −1, 0, and +1.[25] For the control group, $\bar{U}_t$ are slightly higher for the 1970 10-K report as compared to the 1969 10-K report. However, for the experimental group $\bar{U}_t$ for the 1970 10-K reports are substantially higher than those for the 1969 10-K reports. For example, on day −1 $\bar{U}_t$ for the 1969 10-K is 1.428 and for the 1970 10-K it is 1.801. However, for the experimental group there appears to be very little difference between the 1969 and 1970 $\bar{U}_t$ for days +1 and +2. Thus, the empirical evidence for the experimental group appears to indicate that, for days −2, −1, and 0, the price variability around the 1970 10-K report was substantially higher than that of the 1969 10-K report. The last column reports the Z-statistic from the Wilcoxon Signed Ranks Test, a test of equality of 1969 and 1970 U_t (within each group).[26] For the control group there are no statistically significant differences between mean values of U_t for 1969 and 1970. As expected none of the Wilcoxon Z-statistics for the control group is significant at the five percent level (one-tailed). The Wilcoxon Z-statistics for the experimental group are significant at the five percent level or less (one-tailed) on days −2, −1, and 0, but they are not statistically significant at any reasonable level for days +1 and +2. Thus, preliminary evidence presented in table 2 appears to indicate that for the experimental group, the price variability around the dates of release of 10-K reports is significantly higher for the 1970 10-K than for the 1969 10-K. However, the evidence indicates no significant differences in price variability between 1969 and 1970 for the control group.[27]

---

[24] All empirical tests were also conducted on *CHDB* and the results were similar to those of *PCHDB*. *PCHDB* was also measured using $DB_i^{May\ 1971}$ in the denominator, and the results were similar.

The bottom half of table 2 presents preliminary results for divergence of beliefs (DB). For the control group there is little difference in average DB between May 1970 and May 1971 (0.0484 vs. 0.0482). For the experimental group there is a substantial decrease in average DB from May 1970 to May 1971 (0.0508 to 0.0279). The Wilcoxon Z-statistic is equal to −4.89 and is significant at the one percent level (one-tailed).[28] Thus, preliminary evidence appears to indicate that between May 1970 and May 1971, there was a significant decrease in divergence of beliefs for the experimental group, but not for the control group.

---

[25] $\bar{U}_t$ were also examined for days −10 to +10. Except for days −2, −1, 0, and +1 (as reported in table 2) none of the other $\bar{U}_t$ were greater than one. This evidence is consistent with that of Foster and Vickrey (1978), who reported that almost all the capital market reaction to 10-K reports took place in the week of release of the report. However, they used weekly returns, and thus it is not possible to deduce from their study which days in the 10-K report release week the market reaction occurred.

[26] Since the equality of price variability of the same firms in two different time periods is tested, the observations are dependent, and the Wilcoxon Signed Ranks Test is the appropriate test. The interested reader is referred to Berenson and Levine (1979, 422–29) for a description of this test and the derivation of the related Z-statistic. The nonparametric Wilcoxon test is used because for both groups the Kolmogorov-Smirnov (K-S) test rejected at the five percent level or less the hypothesis that the U_t are normally distributed.

[27] Recall that the dates of release of the 1969 and 1970 10-K reports are used as the event dates to measure price variability. For firms in the experimental group, if their 1970 annual reports were released prior to their 10-K reports *and* if their 1970 annual reports contain segment data, then the appropriate event date for those firms should be the dates of release of their annual reports. Sixty-five firms in the experimental group released their 1970 annual reports before their 1970 10-K reports. However, an examination of their annual reports revealed that none of these firms disclosed segment data in its 1970 annual reports. Thus, the firms' first disclosure of segment data was in their 10-K reports, and the date of release of their 10-K reports is the appropriate event date.

[28] Nonparametric test is used because the K-S test rejected normality of DB at the five percent level, for both groups.

Table 3
Intergroup Comparison of Percentage Change in Price Variability and Divergence of Beliefs

Panel A. Percentage Change in Price Variability:[a]

	Control Group[b] (N=101)	Experimental Group[b] (N=160)	
Day	PCHPV$_t$	PCHPV$_t$	*Wilcoxon Z*[c]
−2	0.072	0.298	2.23*
−1	0.064	0.251	2.09*
0	0.057	0.227	1.96*
+1	0.014	0.062	0.61
+2	0.022	0.049	0.57

Panel B. Percentage Change in Divergence of Beliefs:[d]

Control Group (N=101)	Experimental Group (N=160)	
PCHDB	PCHDB	*Wilcoxon Z*[e]
−0.1043	−0.4025	−4.61**

[a] Day 0 is the date of receipt of the 10-K at the SEC. PCHPV$_t$ is the median percentage change in price variability between the 1969 and 1970 10-K reports, on day *t*. *PCHPV$_t$* is truncated at 200 percent.

[b] Control group and experimental group are described in table 1.

[c] Wilcoxon Z is the test statistic from the Wilcoxon Ranks Sum Test of equality of *PCHPV$_t$* between the control and experimental groups. * indicates significance at the five percent level (one-tailed).

[d] *PCHDB* is the median percentage change in divergence of beliefs between May 1970 and May 1971. PCHDB is truncated at 200 percent.

[e] Wilcoxon Z is the test statistic from the Wilcoxon Ranks Sum Test of equality of *PCHDB* between the control and experimental groups. ** indicates significance at the one percent level (one-tailed).

Table 3 presents intergroup comparison results for hypotheses H1 and H2. The top half of the table presents results for hypothesis H1. The median percentage change in price variability on day *t* (*PCHPV$_t$*) is presented for the control and experimental groups.[29] A comparison of *PCHPV$_t$* between the two groups shows that for all five days in the event period, *PCHPV$_t$* of the experimental group is higher than that of the control group. The differences are substantial for days −2, −1, and 0. For days −2, −1, and 0 the Z-statistics from the Wilcoxon Ranks Sum Test are significant at the five percent level (one-tailed).[30] Thus, the results reported in table 3 are consistent with hypothesis H1; the increase in price variability (1969 to 1970 10-K) is significantly higher for the experimental group than for the control group.

---

[29] Some outliers were found in the distribution of PCHPV$_t$. These were set equal to 200 percent. The K-S test rejected normality of *PCHPV$_t$* at the five percent level for all *t*, for both groups. Thus, nonparametric tests are used to test hypothesis H1.

[30] Since the difference in the medians of two groups of firms is tested, the observations are independent, and the Wilcoxon Ranks Sum Test is the appropriate test. The interested reader is referred to Berenson and Levine (1979, 429–35) for a description of this test and the derivation of the related Z-statistic.

The bottom half of table 3 presents results for hypothesis H2. The median percentage change in divergence of beliefs (*PCHDB*) is presented.[31] There is a substantial difference in *PCHDB* between the control group and experimental group (−0.1043 vs. −0.4025). The Wilcoxon Z-statistic is equal to −4.61 and is significant at the one percent level (one-tailed). Thus, the results reported in table 3 are consistent with hypothesis H2; the decrease in divergence of beliefs (May 1970 to May 1971) is significantly greater for the experimental group than for the control group.

Test Method (H3 and H4)

Hypothesis H3 involves testing whether there is a relationship between $PCHPV_t$ and number of segments (*NSEG*).[32] Based on the empirical evidence reported for hypothesis H1, the power of the cross-sectional test is likely to be higher if $PCHPV_t$ for days −2, −1, and 0 are used to formulate the dependent variable for hypothesis H3. For each firm the average percentage change in price variability ($APCHPV_i$) over days −2, −1, and 0 for firm *i* was calculated as follows:

$$APCHPV_i = \frac{1}{3}\sum_{t=-2}^{t=0} PCHPV_{it}. \tag{11}$$

$APCHPV_i$ was regressed on $NSEG_i$ to test hypothesis H3. Similarly, $PCHDB_i$ was regressed on $NSEG_i$ to test hypothesis H4. However, regression diagnostics showed that the error terms for both regressions were not normally distributed as assumed in ordinary least squares regression. A common technique used to mitigate this problem is to use the natural log of the variables (e.g., Brown et al. 1987) after adding positive constants to the dependent variables to make all their values positive. Furthermore, since most of the values of *PCHDB* were negative, to minimize the magnitude of the positive constant to be added, *PCHDB* was first multiplied by −1. The regressions were run using these transformed variables and normality of error terms was no longer rejected. The regressions were run for the control group and experimental group combined, allowing for differences in slope coefficients between the two groups. To test hypothesis H3 the following regression was estimated:

$$LAPCHPV_i = \psi_0 + \psi_1 \cdot LNSEG_i + \psi_2 \cdot (D_{1i} \cdot LNSEG_i) + e_{1i}. \tag{12}$$

$LAPCHPV_i$ and $LNSEG_i$ are the natural log of (the adjusted) $APCHPV_i$ and $NSEG_i$, respectively. D_{1i} is a dummy variable equal to one for the experimental group and zero for the control group, and e_{1i} is the regression error term. ψ_0 is the intercept, ψ_1 is the slope coefficient for the control group, and $\psi_1 + \psi_2$ is the slope coefficient for the experimental group. Since no significant increase in price variability is expected for the control group, no significant relationship is expected between $LAPCHPV_i$ and $LNSEG_i$ for this group. Thus, ψ_1 is hypothesized to be zero. For the experimental group, hypothesis H3 predicts that the more the number of segments the greater the increase in price variability. Thus, $\psi_1 + \psi_2$ is hypothesized to be positive. Similarly, to test hypothesis H4 the following regression was estimated:

$$LPCHDB_i = \lambda_0 + \lambda_1 \cdot LNSEG_i + \lambda_2 \cdot (D_{1i} \cdot LNSEG_i) + e_{2i}. \tag{13}$$

---

[31] Since the K-S test rejected normality of *PCHDB* at the five percent level for both groups, nonparametric tests are used to test hypothesis H2. Outliers in *PCHDB* were set equal to 200 percent.

[32] *NSEG* had a mean of 4.56 for the control group and 3.73 for the experimental group.

$LPCHDB_i$ is the natural log of ($PCHDB_i \cdot -1$). D_{1i} and $LNSEG_i$ are as in equation (12), and e_{2i} is the regression error term. λ_0 is the intercept, λ_1 is the slope coefficient for the control group, and $\lambda_1+\lambda_2$ is the slope coefficient for the experimental group. Since no significant decrease in divergence of beliefs is expected for the control group, no significant relationship is expected between $LPCHDB_i$ and $LNSEG_i$ for this group. Thus, λ_1 is hypothesized to be zero. For the experimental group, hypothesis H4 predicts that the *more* the number of segments, the greater the *decrease* in divergence of beliefs. This implies an inverse relationship between $PCHDB_i$ and $NSEG_i$. However, $PCHDB_i$ was multiplied by -1 before taking the log transformation. Thus, $\lambda_1+\lambda_2$ is hypothesized to be positive.

Results (H3 and H4)

The results of the cross-sectional regressions are presented in table 4.[33] The top half is for regression equation (12), a test of hypothesis H3. As predicted, the slope coefficient for the control group, ψ_1 is not statistically different from zero ($t=0.432$), but the slope coefficient for the experimental group, $\psi_1+\psi_2$ is positive and statistically different from zero at the one percent level (one-tailed $t=2.92$). These results imply that for the experimental group, the magnitude of increase in price variability is positively related to the number of segments, whereas for the control group there is no such relationship. These results are consistent with hypothesis H3.[34] The bottom half of table 4 presents results for regression equation (13), a test of hypothesis H4. The t-statistic on the slope coefficient for the control group (λ_1) is not significant ($t=0.693$), whereas the t-statistic on the slope coefficient for the experimental group ($\lambda_1+\lambda_2$) is positive and statistically significant at the one percent level (one-tailed $t=3.17$). These results imply that the magnitude of decrease in divergence of beliefs is positively related to the number of segments for the experimental, but not for the control group. These results are consistent with hypothesis H4.

Alternative Explanations for the Results

The increase in price variability and decrease in divergence of beliefs documented in this study could be caused by intervening factors other than segment disclosure. For instance, there could be some macroeconomic factors (interest rates, money supply, etc.) that could have changed the price variability and divergence of beliefs. However, there is no a priori reason to expect these factors to affect the control and experimental groups differently. If increase in price variability and decrease in divergence of beliefs

---

[33] An econometric problem frequently encountered in cross-sectional regression is heteroscedasticity of the error terms. This leads to inconsistent estimates of standard error of the coefficients and to faulty inferences when testing hypotheses. A test for heteroscedasticity, developed in White (1980), was conducted on equations (12) and (13). The chi-square statistics from this test were 3.85 ($p=0.57$) and 2.23 ($p=0.81$), respectively. The test did not reject the null hypothesis that the error terms were homoscedastic. Thus, the t-statistics from the ordinary least squares regression will lead to proper inferences.

[34] For completeness, test of hypothesis H3 was also conducted using all five days in the event period (days -2 to $+2$) to calculate $APCHPV$. The result was similar to that reported in table 4. The results were also similar when (the natural log of) $PCHPV_t$ for days -2, -1, and 0 were used individually as the dependent variables in equation (12). However, when (the natural log of) $PCHPV_t$ for days $+1$ and $+2$ were used as the dependent variables the slope coefficients on $LNSEG$ were not statistically different from zero, for control and experimental groups. This is not surprising, since tables 2 and 3 showed that there was no significant increase in price variability in days $+1$ and $+2$, for both groups.

Table 4
Results of Cross-Sectional Regression Tests

Panel A. Price Variability:[a]

Model:[b] $LAPCHPV_i = \psi_0 + \psi_1 \cdot LNSEG_i + \psi_2 \cdot [D_{1i} \cdot LNSEG_i) + e_{1i}$

(N=261) *Coefficients* *(t-statistics)*[c]			
ψ_0	ψ_1	ψ_2	*Adj.* R^2
−1.861 (−2.845)**	0.006 (0.432)	0.070 (2.052)*	0.0264

Test: $\psi_1 + \psi_2 > 0$, $t = 2.92$ (significant at one percent level, one-tailed)

Panel B. Divergence of Beliefs:[d]

Model:[b] $LPCHDB_i = \lambda_0 + \lambda_1 \cdot LNSEG_i + \lambda_2 \cdot (D_{1i} \cdot LNSEG_i) + e_{2i}$

(N=261) *Coefficients* *(t-statistics)*[c]			
λ_0	λ_1	λ_2	*Adj.* R^2
−1.022 (−2.754)**	0.002 (0.693)	0.014 (2.144)*	0.0253

Test: $\lambda_1 + \lambda_2 > 0$, $t = 3.17$ (significant at one percent level, one-tailed)

[a] $LAPCHPV_i$ is Log($APCHPV_i$), where $APCHPV_i$ is average $PCHPV_{it}$ over days −2, −1, and 0, for firm i. Day 0 is the date of receipt of 10-K at the SEC. $PCHPV_{it}$ is the percentage change in price variability from 1969 to 1970 10-K report for firm i on day t.

[b] $LNSEG_i$ is Log($NSEG_i$) where $NSEG_i$ is the number of segments disclosed by firm i. D_{1i} is a dummy variable equal to one for the experimental group and equal to zero for the control group. Control and experimental groups are described in table 1.

[c] *(**) indicates significance at the five percent (1 percent) level. The significance levels are one-tailed for ψ_1, ψ_2, λ_1, and λ_2, and two-tailed for ψ_0 and λ_0.

[d] $LPCHDB_i$ is Log[($PCHDB_i$)(−1)] where $PCHDB_i$ is the percentage change in divergence of beliefs from May 1970 to May 1971, for firm i.

is due to macroeconomic factors, it is unlikely that it would be correlated with the number of segments.[35]

Increase in price variability and decrease in divergence of beliefs for the experimental group is attributed to the increase in the precision of information due to disclosure of SEC mandated segment data. If this relationship is valid, then for the experimental group there should be a negative correlation between change in price variability and change in divergence of beliefs. For the experimental group the correlation between *APCHPV* and *PCHDB* is −0.71 (statistically significant at the one percent level), which gives greater credence to the hypotheses.

---

[35] If the number of segments is related to firm size and if increase in price variability and decrease in divergence of beliefs is related to firm size, then number of segments could be a proxy for firm size in the results reported in table 4. Using the market value of equity as of 31 December 1970 as a measure of firm size, it was found that, for the control group and the experimental group, there was a statistically significant positive correlation between firm size and number of segments (correlation of 0.22 and 0.24, respectively, both significant at the five percent level). However, regression analysis similar to that in table 4, showed that for both the control and experimental groups, there was no statistically significant relationship between firm size and increase in price variability or between firm size and decrease in divergence of beliefs. Thus, it is very unlikely that in the cross-sectional tests, number of segments is acting as a proxy for firm size.

Another possibility is the results could be due to other accounting regulations coming into effect at the same time as the SEC segment regulation. An examination of accounting regulations during that time period revealed that Accounting Principles Board Opinion No. 16 (APB 16) on business combinations issued in August 1970 was the only regulation that could have had an immediate impact on the 1970 10-K reports of the sample firms. APB 16 laid down the criteria under which a business combination can be accounted for under the pooling of interests method. Business combinations that did not meet all of the specified criteria had to be accounted for under the purchase method. APB 16 was applicable to all business combinations entered into after 31 October 1970. Thus, if any of the firms in the sample entered into a business combination between 31 October 1970 and 31 December 1970, their 1970 10-K reports would be affected. However, a careful examination of the 10-K and annual reports, proxy, and registration statements for the relevant time period showed that none of the firms in the sample engaged in any business combinations during this time period. Thus, APB 16 is not expected to have affected the firms in the sample.

Finally, the price variability results could be due to differences in the timing of release of 10-K reports vis-à-vis annual reports. If the number of experimental group firms releasing their annual reports before their 10-K reports was lower for fiscal year 1970 than for fiscal year 1969, then there is a possibility that nonsegment information provided to the market in the 1970 10-K is greater than that provided in the 1969 10-K. This could result in price variability for the 1970 10-K being higher than that for the 1969 10-K. To explore this possibility, the number of firms releasing their annual reports before their 10-K was examined for both groups, for both fiscal years. For fiscal year 1969, 54 firms in the control group and 62 firms in the experimental group released their annual reports before their 10-K reports. For fiscal year 1970 the numbers were 57 and 65, respectively. Thus, the number of firms releasing annual reports before 10-K reports was almost the same for fiscal years 1969 and 1970, and it is unlikely that the price variability results are being driven by timing differences.[36]

IV. Summary and Conclusions

The purpose of this study was to provide further empirical evidence on the economic impact of SEC mandated segment data. Methodological deficiencies in the previous security price studies were overcome by identifying the exact date of release of 10-K reports and by using the price variability methodology. The previous earnings predictability studies were extended by examining segment data's impact on divergence of beliefs as measured by the coefficient of variation of multiple analysts' earnings fore-

---

[36] The possibility of testing hypothesis H1 separately, for those firms releasing their annual report before their 10-K, and for those firms releasing their 10-K before their annual report, was considered. However, this test can be conducted only on those firms for which the annual report is released before the 10-K (or vice versa) both for fiscal 1969 *and* 1970. For the control group (experimental group), 19 (four) firms released their annual report before their 10-K for both fiscal 1969 and 1970, and nine (37) firms released their 10-K before their annual report for both fiscal 1969 and 1970. Statistical tests conducted on such small sample sizes are likely to be unreliable.

casts. The empirical evidence showed that both the increase in price variability and decrease in divergence of beliefs were significantly more for the experimental than for the control group. Furthermore, for the experimental group, the magnitudes of increase in price variability and decrease in divergence of beliefs were directly proportional to the number of segments, whereas no such relationship was documented for the control group. In terms of the various economic contexts in which the impact of SEC mandated segment data has been studied so far, the empirical evidence shows that segment data were associated with a reduction in the systematic risk, increase in the accuracy of earnings forecasts, a lower divergence of beliefs, and an increase in price variability around the dates of release of 10-K reports containing segment data. The first two effects have been documented by previous research, and the last two effects were documented in this study.

Appendix

Frequency Distribution of 10-K Report Filing Dates

	1969 10-K Reports				*1970 10-K Reports*			
	Control Group[a]		*Experimental Group*[a]		*Control Group*		*Experimental Group*	
Period[b]	*Frequency*	*%*	*Frequency*	*%*	*Frequency*	*%*	*Frequency*	*%*
2/1–2/10	3	3.0	1	0.6	4	4.0	0	0.0
2/11–2/20	6	5.9	12	7.5	5	5.0	5	3.2
2/21–2/28	14	13.9	23	14.4	21	20.7	29	18.1
3/1–3/10	22	21.8	34	21.2	17	16.8	28	17.5
3/11–3/20	21	20.8	35	21.9	20	19.7	35	21.9
3/21–3/31	28	27.6	48	30.0	24	23.8	47	29.3
4/1–4/10	2	2.0	3	1.9	4	4.0	9	5.6
4/11–4/20	4	4.0	3	1.9	3	3.0	5	3.2
4/21–4/30	1	1.0	1	0.6	3	3.0	2	1.2
Total	101	100.0	160	100.0	101	100.0	160	100.0

[a] Control group and experimental group are described in table 1.

[b] Fiscal year 1969 (1970) 10-K reports are released in calendar year 1970 (1971).

References

AICPA (American Institute of Certified Public Accountants). 1970. *Accounting Principles Board, Opinion No. 16: Business Combinations.* New York: AICPA.

Ajinkya, B. 1980. An empirical evaluation of line of business reporting. *Journal of Accounting Research* 18 (Autumn): 343–61.

———, and M. Gift. 1985. Dispersion of financial analysts' earnings forecasts and the (option model) implied standard deviations of stock returns. *Journal of Finance* 40 (December): 1353–65.

Baldwin, B. 1984. Segment earnings disclosure and the ability of security analysts to forecast earnings per share. *The Accounting Review* 59 (July): 376–89.

Barefield, R., and E. Comiskey. 1975. The impact of the SEC's line of business disclosure on the accuracy of analysts' forecasts of earnings per share. Working paper, University of Arizona, Tucson.

Barry, C., and S. Brown. 1985. Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis* 20 (December): 407–22.

Berenson, M., and D. Levine. 1979. *Basic Business Statistics, Concepts and Applications.* Englewood Cliffs, NJ: Prentice-Hall.

Bhushan, R. 1989. Collection of information about publicly traded firms: Theory and evidence. *Journal of Accounting & Economics* 11 (July): 183–206.

Brown, L., G. Richardson, and S. Schwager. 1987. An information interpretation of financial analyst superiority in forecasting earnings. *Journal of Accounting Research* 25 (Spring): 49–67.

Choi, S., and G. Salamon. 1989. External reporting and capital asset prices. Working paper, Vanderbilt University, Nashville, TN.

Collins, D. 1976. Predicting earnings with subentity data: Some further evidence. *Journal of Accounting Research* 14 (Spring): 163–77.

———, and R. Simonds. 1979. SEC line of business disclosure and market risk adjustments. *Journal of Accounting Research* 17 (Autumn): 352–83.

Cukierman, A., and D. Givoly. 1982. Heterogeneous earnings expectations and earnings uncertainty: Theory and evidence. Working paper, Tel Aviv University, Israel.

———, and P. Wachtel. 1979. Differential inflationary expectations and the variability of the rate of inflation: Theory and evidence. *American Economic Review* 69 (September): 595–609.

Daley, L., D. Senkow, and R. Vigeland. 1988. Analysts' forecasts, earnings variability, and option pricing: Empirical evidence. *The Accounting Review* 63 (October): 563–85.

Elton, E., M. Gruber, and M. Gultekin. 1984. Professional expectations: Accuracy and diagnosis of errors. *Journal of Financial and Quantitative Analysis* 19 (December): 351–63.

FASB (Financial Accounting Standards Board). 1976. *Statement of Financial Accounting Standards No. 14: Financial Reporting for Segments of a Business Enterprise.* Norwalk, CT: FASB.

Foster, T., and D. Vickrey. 1978. The incremental information content of the 10-K. *The Accounting Review* 53 (October): 921–34.

Holthausen, R., and R. Verrecchia. 1981. The change in price variability resulting from a sequence of information releases. Working paper, University of Chicago.

———. 1988. The effect of sequential information releases on the variance of price changes in an intertemporal multi-asset market. *Journal of Accounting Research* 26 (Spring): 82–106.

———. 1990. The effect of informedness and consensus on price and volume behavior. *The Accounting Review* 65 (January): 191–208.

Horwitz, B., and R. Kolodny. 1977. Line of business reporting and security prices: An analysis of an SEC disclosure rule. *Bell Journal of Economics* 8 (Spring): 234–49.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47 (December): 587–616.

Mohr, R. 1983. The segmental reporting issue: A review of empirical research. *Journal of Accounting Literature* 2 (Spring): 39–72.

Ohlson, J. 1979. On financial disclosure and the behavior of security prices. *Journal of Accounting & Economics* 1 (December): 211–32.

Patell, J. 1976. Corporate forecasts of earnings per share and stock price behavior: Empirical tests. *Journal of Accounting Research* 14 (Autumn): 246–76.

SEC (Securities and Exchange Commission). 1966. *Testimony of Chairman Cohen During Congressional Hearings on Economic Concentration.* Washington, D.C.: SEC.

———. 1969. *Securities Act of 1933, Release No. 4949.* Chicago: Commerce Clearing House.

———. 1970. *Securities Exchange Act of 1934, Release No. 9000.* Chicago: Commerce Clearing House.

Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19 (September): 425–42.

Simonds, R., and D. Collins. 1978. Line of business reporting and security prices: An analysis of an SEC disclosure rule: Comment. *Bell Journal of Economics* 9 (Autumn): 646–58.

White, H. 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* 48 (May): 817–38.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 42–55

The Market Reaction to 10-K and 10-Q Filings and to Subsequent *The Wall Street Journal* Earnings Announcements

Earl K. Stice
Rice University

SYNOPSIS: Firms sometimes file their 10-K or 10-Q with the Securities and Exchange Commission (SEC) several days before the corresponding earnings announcement appears in *The Wall Street Journal* (WSJ). In these cases, the 10-K or 10-Q filing constitutes the first public announcement of earnings. This study addresses the question of whether the price and volume reactions to these earnings announcements occur at the SEC filing date or at the subsequent WSJ announcement date. This research can be viewed as a limited test of whether the securities markets are efficient with respect to the data contained in 10-K and 10-Q filings. The generalizability of the test results is limited because of the nonrandom nature of the sample.

Tests were performed on daily price and volume data for 342 firm-quarters for which the SEC filing date preceded the WSJ earnings announcement by at least four trading days. The results suggest that there was no significant market reaction, on average, at the SEC filing date, even though the filing was the first public announcement of earnings for the quarter. However, there is evidence of the existence of a market reaction to the subsequent WSJ earnings announcement. The direction of the price reaction at the WSJ announcement date is consistent with the sign of unexpected earnings.

Wright and Groff (1986) present evidence indicating that the existence

---

This paper is based on my doctoral dissertation at Cornell University. The help of my committee members—Dale Morse (chairman), Tom Dyckman, and Maureen O'Hara—is gratefully acknowledged. Thanks also to the Deloitte, Haskins & Sells Foundation for financial support and to Disclosure Information Group for providing data. In addition, the helpful comments of workshop participants at the University of Arizona, the editor, and the anonymous referees are greatly appreciated.

Submitted January 1989.
Revisions received July 1989 and May 1990.
Accepted September 1990.

of a market reaction to the public release of price-relevant information does not depend on exactly how that information is made public. In contrast, the results presented in this paper suggest that, in some limited cases, the method of disclosing accounting earnings is related to whether the information embodied in those earnings will be reflected in security prices in a timely fashion. The results also suggest that, at least in one specific set of circumstances, data disclosed as part of an SEC-mandated filing are not fully reflected in prices until a subsequent media disclosure is made.

Key Words: *10-K/10-Q filings, Earnings announcements, Market efficiency.*

Data Availability: *A list of the sample firms, the event dates, and the price and volume data used in this study are available from the author.*

EXTENSIVE evidence has been accumulated demonstrating that earnings announcements appearing in *The Wall Street Journal* (WSJ) are associated with a significant increase in price and volume activity in the trading days surrounding the announcement (Beaver 1968; Bamber 1986; Hughes and Ricks 1987). The WSJ earnings announcement is typically associated with the first public disclosure of earnings. However, in some cases, firms file their 10-K or 10-Q with the Securities and Exchange Commission (SEC) several days before the earnings contained in the filing are announced in the WSJ. These firms are used in this study to examine whether the information content[1] of an earnings announcement can be affected by the method in which the earnings are announced.

Beaver (1981) defines a security market as being efficient with respect to a specific information set if the prices that exist in that market are the same as the prices that would exist if every individual trader had that information set. For the sample firms, the existence of a market reaction to the WSJ announcements but not to the preceding SEC filings is consistent with the interpretation that, at least in this very limited setting, the market is not efficient with respect to financial data filed with the SEC.

This paper is organized as follows: Section I is a description of the sample identification process. Section II contains the results of tests for a market reaction to the SEC filings and to the subsequent WSJ earnings announcements, and section III is a summary and conclusion.

I. Sample Selection and Data

The sample was selected to include only firm-quarters for which the receipt date of the 10-K or 10-Q filing at the Securities and Exchange Commission preceded the corresponding earnings announcement in the WSJ by at least four trading days. Initially, a set of candidate sample firm-quarters was machine generated by comparing the COMPUSTAT earnings report date (a proxy for the WSJ earnings announcement date) and the date the 10-K or 10-Q was received by Disclosure, Inc. (a proxy for the SEC receipt date).[2] Firm-quarters were considered candidates for the sample if the Disclosure, Inc., receipt date preceded the COMPUSTAT report date by at least three trading days. In all, 677 firm-quarters from March 1976 through September 1985 were identified.

---

[1] The term *information content* of an accounting announcement is used to refer to the extent to which the announcement results in changes in equilibrium prices or portfolio choice. Accordingly, the focus is on short-run (daily) price and volume reactions to accounting announcements.

For each of these 677 firm-quarters, the WSJ Index was searched for the actual WSJ earnings announcement date. Of the 520 WSJ Index entries found, only 297 (57 percent) corresponded with the earnings report date on COMPUSTAT. Hughes and Ricks (1987) report that with their sample of annual earnings announcement dates, the COMPUSTAT date and the WSJ Index date were the same in 75 percent of the cases. The lower proportion here is due to the presence of a number of COMPUSTAT errors among the 677 firm-quarters. For example, for 75 firm-quarters, the COMPUSTAT report date followed the WSJ announcement by more than a month (22 trading days or more). The proportion of COMPUSTAT errors among the 677 candidate firm-quarters is high because the SEC filing date is more likely to precede the reported COMPUSTAT earnings announcement date if that announcement date is in error. Accordingly, all dates were manually verified.

The actual WSJ earnings announcement date was compared to the Disclosure, Inc., receipt date for the 520 firm-quarters for which a WSJ Index entry was located. The Disclosure, Inc., receipt date preceded the WSJ earnings announcement date by at least three trading days for 368 firm-quarters (representing 197 firms). For these 368 firm-quarters, the actual SEC receipt date was obtained from the SEC on-line data base or from microfiche copies of the filings. SEC receipt dates were found for 360 of the 368 firm-quarters, with 348 of these preceding the WSJ earnings announcement by at least four trading days.

In order to broaden the sample, a random sample of 4,200 Disclosure receipt dates for filings from non-COMPUSTAT firms was generated from the Disclosure, Inc., tape, and the WSJ Index was searched for the associated earnings announcement dates. Of the 1,441 WSJ earnings announcements found, 71 were preceded by the receipt of the 10-K or 10-Q filing at Disclosure, Inc., by at least four calendar days. For 65 of these firm-quarters the actual SEC receipt date preceded the WSJ date by at least four trading days.

In total, then, 413 firm-quarters (representing 240 firms) were identified for which the SEC receipt date preceded the WSJ earnings announcement date by at least four trading days—348 COMPUSTAT firm-quarters (representing 190 firms), and 65 non-COMPUSTAT firm-quarters (representing 50 firms).

In order to meet the statutory filing time requirement, firms sometimes file a 10-K or 10-Q without including financial statements, which would be filed later as an amendment to the original filing. Fifteen cases were identified in which the original filing did not contain financial statements, and for this reason they were dropped from the sample.

The SEC filing preceded the WSJ announcement because the WSJ announcement was late, not because the SEC filing was early. In fact, there is little dispersion at all in the SEC filing lag periods; about 75 percent of the 10-K and 10-Q filings for the sample firm-quarters occurred within one business day of the statutory limit (45 days for 10-Q filings and 90 days for 10-K filings). The lag between fiscal quarter end and the WSJ earnings announcement increased ten calendar days (median increase) in the sample firm-quarters compared to the quarters in the preceding year. In absolute terms, the median lags between fiscal quarter end and the WSJ announcement were 57 days for quarterly announcements and 102 days for annual announcements. Compared to the WSJ announcement lags reported by Chambers and Penman (1984) of 25 days for quarterly announcements and 44 days for annual announcements, the WSJ announcements in the sample firm-quarters were significantly delayed.

---

[2] The Disclosure Information Group provided a magnetic tape containing 10-K and 10-Q receipt dates used in this research.

As an attempt to identify the predominant cause for the earnings announcement delays, press release dates for each firm-quarter were sought on P.R. Newswire, a financial news release wire service. Access to past P.R. Newswire releases, starting from 21 January 1980, is available through NEXIS. Of the 310 sample earnings announcements that fell within the P.R. Newswire coverage period, only 88 (28 percent) were represented by a press release in the P.R. Newswire file. For these 88 press releases, 52 (59 percent) occurred the day before or the day of the WSJ announcement, 33 (38 percent) were made at least two business days before the announcement, and 3 were made after the WSJ announcement. Both firm delays and WSJ delays appeared to be responsible for the late appearance of the sample firm earnings announcements. Since exchange rules require timely public release of quarterly and annual earnings, NYSE and AMEX personnel were contacted to determine what exchange-imposed penalties, if any, a firm could expect for making late financial statement disclosures. Both exchanges indicated that no routine penalties are imposed for marginally late disclosures. In extreme cases, failure to make financial statement disclosures can result in trading halts or delisting. No such sanctions were imposed on any of the sample firms as a result of the delayed earnings announcements.

A financial profile of the sample firms indicates that, on average, they are small firms experiencng financial difficulties for the quarters appearing in the sample. Losses were reported in more than 40 percent of the sample firm-quarters, and market values and quarterly sales decreased compared to the prior year. Median quarterly revenues for the sample firm-quarters were $13.3 million (in 1988 dollars), compared to $94.0 million for the population of COMPUSTAT firms from the same time period. Median market values were $32.4 million (in 1988 dollars), compared to $196.1 million for all COMPUSTAT firms. Also, the sample firms were not widely followed by analysts—analysts' earnings forecasts made in the three months preceding the SEC filing were available in the I/B/E/S database for only 106 of the 398 sample firm-quarters.

For purposes of the market reaction tests reported below, firm-quarters were eliminated from the sample if the WSJ Index indicated the announcement of any other significant event (i.e., stock split, merger, dividend increase, new common stock issue, etc.) within two weeks of either the SEC or WSJ date. This resulted in the elimination of 43 firm-quarters.

Daily stock returns were obtained for the 60-day period surrounding the SEC filing and the WSJ announcement (SEC_{-27} through SEC_{+2} and WSJ_{-1} through WSJ_{+28}) for 342 of the remaining 355 firm-quarters. Returns were extracted from the CRSP daily returns file when available. Otherwise, returns were manually obtained from Standard & Poor's *Daily Stock Price Record*. For OTC firms, returns were computed as the percentage

Table 1
Sample Selection

	Firm-Qtrs.	Firms
COMPUSTAT Firms:		
All firm-quarters between March 1976 and September 1985 for which earnings announcement dates were available on the 1986 quarterly COMPUSTAT tape.	68,479	2,318
All of the above with corresponding 10-K or 10-Q receipt dates available on the Disclosure tape.	60,090	2,097
Disclosure receipt date preceded the COMPUSTAT earnings announcement date by three or more trading days.	677	353
WSJ Index earnings announcement located.	520	306
Disclosure receipt date preceded the *WSJ Index* earnings announcement date by three or more trading days.	368	197
SEC receipt date located.	360	195
SEC receipt date preceded the *WSJ Index* earnings announcement date by four or more trading days.	348	190
Filing contained financial statements.	336	184
No confounding events within two weeks of SEC or *WSJ* date.	297	169
Daily returns available for the entire 60-day test period.	287	165
Non-COMPUSTAT Firms:		
Disclosure receipt date preceded the *WSJ Index* earnings announcement date by four or more calendar days.	71	55
SEC receipt date located.	70	54
SEC receipt date preceded the *WSJ Index* earnings announcement date by four or more trading days.	65	50
Filing contained financial statements.	62	48
No confounding events within two weeks of SEC or *WSJ* date.	58	47
Daily returns available for the entire 60-day test period.	55	46

change in the average of the closing bid and ask prices. Daily trading volume for 339 firm-quarters was also obtained from the *Daily Stock Price Record*.

The sample selection process is summarized in table 1.

Mean-adjusted returns are used as a measure of excess returns. The standardized mean-adjusted return is defined as follows:

$$r_{it} \equiv R_{it} - \bar{R}_{i}, \tag{1}$$

where:

$R_{it} \equiv$ return for firm-quarter i for three-day interval t of the test period, and
$\bar{R}_{i} \equiv$ mean return for firm-quarter i for the 20 three-day intervals in the 60-day test period.

The values of r_{it} are assumed to be normally distributed with mean zero and variance σ^2_{Ri}. A standardized residual is defined as:

$$SR_{it} \equiv \frac{r_{it}}{(s^2_{Ri})^{1/2}} \bullet [(T_i - 3)/(T_i - 1)]^{1/2}, \tag{2}$$

where:

$s_{Ri}^2 \equiv$ the sample variance of returns for firm-quarter i for the 20 three-day intervals in the 60-day test period, and

$T_i \equiv$ the number of return intervals in the test period for firm-quarter i. $T_i = 20$ for all i.

It can be shown that SR_{it} has an expected value of zero and a variance of one and that $(SR_{it})^2$ has an expected value of one and a variance of $2[(T_i-2)/(T_i-5)]$.[3] The measure $(SR_{it})^2$ is an indication of the magnitude of the variance of excess returns in the interval t, relative to the variance in the entire 60-day period. Similar measures frequently have been used to indicate the existence of abnormal price activity (Beaver 1968; Lobo and Mahmoud 1989).

The following statistic has an asymptotic standard normal distribution:

$$Z_{SR_t} \equiv \frac{\frac{1}{N}\sum_{i=1}^{N}(SR_{it}^2-1)}{[2(T_i-2)/N(T_i-5)]^{1/2}}, \tag{3}$$

where N is the number of firm-quarters in the sample. This Z_{SR_t}-statistic is used to test for the significance of abnormal price activity around the SEC and WSJ dates.

Abnormal trading volume is also used as an indication of information content. Following prior research (Morse 1981; Bamber 1986), percentage of shares traded is used as the measure of trading volume. The standardized excess volume measure is:

$$SV_{it} \equiv \frac{V_{it}-\bar{V}_i}{(s_{Vi}^2)^{1/2}} \cdot [(T_i-3)/(T_i-1)]^{1/2}, \tag{4}$$

where:

$V_{it} \equiv$ total trading volume for firm-quarter i for three-day interval t of the test period, divided by number of shares outstanding.

$\bar{V}_i$ and s_{Vi}^2 are defined analogously to the corresponding return measures. If excess volume $(V_{it}-\bar{V}_i)$ is assumed to have a normal distribution with a mean of zero, then SV_{it} has a mean of zero and a variance of one. In addition, the following statistic has an asymptotic standard normal distribution:

$$Z_{SV_t} \equiv \frac{\frac{1}{N}\sum_{i=1}^{N} SV_{it}}{(1/N)^{1/2}}. \tag{5}$$

The statistic Z_{SV_t} is used to test for the existence of significant increases in trading volume around the SEC and WSJ dates.

---

[3] This expression for the variance differs from that in Lobo and Mahmoud (1989) because of a difference in degrees of freedom—one parameter was estimated (the mean), rather than two as in the market model used by Lobo and Mahmoud. Another difference here is that the parameters were estimated using test period data, whereas Lobo and Mahmoud used a separate estimation period. Also, cross-sectional dependence has been ignored here.

II. Market Reactions at the SEC and WSJ Dates

For the sample firm-quarters, a market response might be expected at the SEC filing date since the filing constitutes the initial public announcement of earnings. In fact, these SEC filings might be more likely to have information content because of the large amount of data disclosed in a 10-K or 10-Q filing, compared to just the EPS and sales figures disclosed in a typical WSJ Earnings Digest announcement. Using 10-K filing and annual report release dates, Foster and Vickrey (1978) and Wilson (1987) demonstrate the incremental information content of these additional 10-K and annual report disclosures beyond earnings and sales.

On the other hand, the extent of dissemination of the SEC filings for the sample firm-quarters is more limited than for a standard WSJ earnings announcement. This limited dissemination of 10-K and 10-Q filings during the first few days after the filings are made is partially a result of the relatively high cost (compared to the cost of obtaining a WSJ earnings announcement) involved in obtaining SEC filings in a timely fashion.[4] To the extent that this limited dissemination impedes the reflecting of the filing data into security prices, there may not be a significant market reaction at the SEC filing date.

For the sample firm-quarters, when the WSJ earnings announcement is made, all financial data contained therein have been made public previously in the 10-K or 10-Q filing. It might be expected that no market reaction would occur as a result of the redundant publication of earnings and sales. However, if the limited dissemination of the SEC filing impedes the incorporation of the data contained therein into prices, then there may be a subsequent market reaction at the WSJ announcement date as the data become more widely disseminated.

Both security returns and trading volume are used to test for the existence of a market reaction. Table 2, panel A, contains the squared return residuals $[(SR_{it})^2]$ for the 20 three-day intervals in the test period. The average squared residual, across all firm-quarters for all 20 three-day intervals, is 0.850. The average for the SEC interval (defined as SEC_0, SEC_{+1}, and SEC_{+2}) is 0.755, suggesting no increase in return variance around the SEC filing date. The average squared residual for the WSJ interval (defined as WSJ_{-1}, WSJ_0, and WSJ_{+1}) is 1.101, indicating an increase in return variance around the WSJ earnings announcement. However, the Z_{SR_t} value for the WSJ interval is only 1.208, suggesting that the increased return variance is significant at only the 0.114 level (one-tailed).

The Z_{SR_t} column of table 2, panel A, reveals that, although the value for the WSJ interval is not significant when compared to the standard normal distribution, it is the greatest of the values for the 20 intervals. In addition, the values seem negatively skewed, and, in fact, 11 of the 20 are less than -2.00. Prior research has suggested that analysis of squared return residuals using the standard normal distribution may not be appropriate. For example, Lobo and Mahmoud (1989) show that the distribution of

---

[4] For individuals outside the filing firms, 10-K and 10-Q filings can be acquired on a timely basis by copying them at the SEC public reference room in Washington or by hiring an information service to do so. For example, Disclosure, Inc., offers an SEC Watch Service in which they search SEC filings for any item which a customer prespecifies and automatically deliver the filings within 24 hours. Judging from the fact that relatively few investors use this type of service to get immediate access to 10-K and 10-Q filings, investors must view the costs as, on average, outweighing the potential benefits.

Table 2

Tests of Squared Standardized Return Residuals from the 60-Day Period Surrounding the SEC Filings and the *WSJ* Earnings Announcements

Panel A. Test Results Based on 342 Sample Firm-Quarters:

Three-Day Interval	*Average* $(SR_{it})^2$	Z_{SR_t}	Pr(Z_{SR_t} ≤ expected value) *NT*	 *ED*
SEC_{-27}–SEC_{-25}	0.943	−0.686	0.754	0.139
SEC_{-24}–SEC_{-22}	0.781	−2.619	0.996	0.802
SEC_{-21}–SEC_{-19}	0.923	−0.922	0.821	0.195
SEC_{-18}–SEC_{-16}	0.824	−2.107	0.982	0.625
SEC_{-15}–SEC_{-13}	0.947	−0.631	0.736	0.130
SEC_{-12}–SEC_{-10}	0.800	−2.382	0.991	0.725
SEC_{-9}–SEC_{-7}	0.729	−3.231	0.999	0.948
SEC_{-6}–SEC_{-4}	0.769	−2.757	0.997	0.850
SEC_{-3}–SEC_{-1}	0.826	−2.074	0.981	0.609
SEC_{0}–SEC_{+2}	0.755	−2.923	0.998	0.890
WSJ_{-1}–WSJ_{+1}	1.101	1.208	0.114	0.002
WSJ_{+2}–WSJ_{+4}	0.985	−0.178	0.571	0.058
WSJ_{+5}–WSJ_{+7}	0.835	−1.966	0.975	0.563
WSJ_{+8}–WSJ_{+10}	0.818	−2.178	0.985	0.646
WSJ_{+11}–WSJ_{+13}	0.861	−1.660	0.952	0.442
WSJ_{+14}–WSJ_{+16}	0.844	−1.860	0.969	0.524
WSJ_{+17}–WSJ_{+19}	0.799	−2.394	0.992	0.729
WSJ_{+20}–WSJ_{+22}	0.816	−2.195	0.986	0.651
WSJ_{+23}–WSJ_{+25}	0.876	−1.475	0.930	0.372
WSJ_{+26}–WSJ_{+28}	0.767	−2.780	0.997	0.851

Panel B. Empirical Distribution of Z_{SR_t} Based on 999 Simulated Observations:

	Standard Normal	*Empirical*
Mean	0.000	−1.767
Standard Deviation	1.000	0.956
1st percentile	−2.326	−3.665
5th percentile	−1.645	−3.257
25th percentile	−0.674	−2.446
50th percentile	0.000	−1.811
75th percentile	0.674	−1.137
95th percentile	1.645	−0.099
99th percentile	2.326	0.688

Average $(SR_{it})^2$ = cross-sectional average of the squared standardized residuals
Z_{SR_t} = statistic generated from $(SR_{it})^2$ by standardizing by the theoretical mean and standard deviation
NT = significance level of Z_{SR_t} based on the standard normal distribution
ED = significance level of Z_{SR_t} based on the empirical distribution

Z-statistics derived from squared residuals has a lower mean than predicted by normal distribution theory. Accordingly, following Lobo and Mahmoud, the significance of the Z_{SR_t}s is evaluated by comparison to an empirically generated distribution as described in Noreen (1989). The empirical distribution is generated in the following way:

1. For each firm-quarter i, a random sample of size 20 is selected, with replacement, from the set of 20 three-day returns.
2. This sample is used to compute a new mean and variance for the returns.
3. Using the new mean and variance along with the random sample of returns, a new set of 20 standardized residuals (SR_{it}) is generated for each firm-quarter.
4. One of the new standardized residuals is randomly selected from each firm-quarter, resulting in a new set of 342 SR_{it}s. A value for Z_{SR_t} is computed using this set of 342 SR_{it}s.
5. Steps 1 through 4 are repeated 999 times, generating an empirical distribution for Z_{SR_t}.
6. The distribution of the 999 values of Z_{SR_t} is used to determine the significance levels of the actual Z_{SR_t} values for the test period. If an observed value for Z_{SR_t} is greater than, say, 950 of the 999 generated values, then the observed value is said to be significant at the 0.05 level.

These results are also reported in table 2, panel A. The descriptive statistics of the empirical distribution are summarized in table 2, panel B. The mean of the empirical distribution, −1.767, is comparable to that reported by Lobo and Mahmoud (1989, 125).[5] Using the empirical distribution, the price reaction during the WSJ interval is shown to be significant at the 0.002 level; that is, of the 999 randomly generated values of Z_{SR_t}, only one was greater than 1.208, the observed Z_{SR_t} for the WSJ interval.

In addition, another distribution was generated using out-of-sample data. To simplify data gathering, only the 230 sample firm-quarters with daily returns available on CRSP were used. The out-of-sample distribution was generated using the same steps outlined above with the exception that the random set of 20 three-day returns for each firm-quarter in each iteration was drawn from the 60-day period preceding the test period. For these 230 firm-quarters, the Z_{SR_t} value for the WSJ interval is 0.530, which, when compared to the out-of-sample empirical distribution, is significant at the 0.025 level.

The market reaction results using trading volume are reported in table 3, panel A. For the entire 60-day test period, the cross-sectional average percentage of shares traded during a three-day interval was 0.419 percent; the average percentages were 0.377 for the SEC interval and 0.504 for the WSJ interval. The values for Z_{SV_t} indicate that there is no evidence of significantly increased trading volume in any interval except for the WSJ interval, where the increase in trading volume is highly significant ($p<0.001$) when evaluated using the standard normal distribution. As with the returns tests, the significance of the volume tests is also determined using an empirical distribution generated in a similar manner. These results are also presented in table 3, panel A; the empirical distribution is summarized in panel B. As with the standard normal results, the statistical conclusions based on the empirical distribution suggest that there is

---

[5] The standard deviation of the empirical distribution reported in table 2, panel B (0.956), is smaller than those reported by Lobo and Mahmoud (1989, 125), which range from 1.317 for large firms to 1.827 for small firms. To see if this difference is a result of the difference in return standardization (see fn. 3), an empirical distribution was generated using the Lobo and Mahmoud standardization on a subset of the sample. Daily returns for the 230 sample firm-quarters available on CRSP were used to generate market-model residuals where the market model was estimated over the 200-day period preceding the test period. The standard deviation of the empirical distribution of Z_{SR_t} generated using these residuals was 1.592.

Table 3
Tests of Trading Volume from the 60-Day Period Surrounding the SEC Filings and the WSJ Earnings Announcements

Panel A. Test Results Based on 339 Sample Firm-Quarters:

Three-Day Interval	*Average* V_{it}	Z_{SV_t}	$\Pr(Z_{SV_t} \leq$ expected value) *NT*	*ED*
SEC_{-27}–SEC_{-25}	0.00420	1.387	0.083	0.077
SEC_{-24}–SEC_{-22}	0.00412	0.223	0.412	0.395
SEC_{-21}–SEC_{-19}	0.00429	1.334	0.091	0.080
SEC_{-18}–SEC_{-16}	0.00373	−1.642	0.950	0.958
SEC_{-15}–SEC_{-13}	0.00419	0.298	0.383	0.372
SEC_{-12}–SEC_{-10}	0.00386	−1.206	0.886	0.902
SEC_{-9}–SEC_{-7}	0.00415	−1.025	0.847	0.852
SEC_{-6}–SEC_{-4}	0.00412	−0.850	0.802	0.793
SEC_{-3}–SEC_{-1}	0.00404	−0.397	0.654	0.651
SEC_{0}–SEC_{+2}	0.00377	−0.973	0.835	0.836
WSJ_{-1}–WSJ_{+1}	0.00504	4.132	0.000	0.001
WSJ_{+2}–WSJ_{+4}	0.00437	0.939	0.174	0.163
WSJ_{+5}–WSJ_{+7}	0.00477	1.142	0.127	0.111
WSJ_{+8}–WSJ_{+10}	0.00422	0.280	0.390	0.377
WSJ_{+11}–WSJ_{+13}	0.00405	−1.285	0.901	0.918
WSJ_{+14}–WSJ_{+16}	0.00408	−0.299	0.618	0.620
WSJ_{+17}–WSJ_{+19}	0.00412	0.089	0.465	0.457
WSJ_{+20}–WSJ_{+22}	0.00426	−0.943	0.827	0.825
WSJ_{+23}–WSJ_{+25}	0.00434	0.042	0.483	0.477
WSJ_{+26}–WSJ_{+28}	0.00402	−1.244	0.893	0.912

Panel B. Empirical Distribution of Z_{SV_t} Based on 999 Simulated Observations:

	Standard Normal	*Empirical*
Mean	0.000	−0.010
Standard Deviation	1.000	0.967
1st percentile	−2.326	−2.170
5th percentile	−1.645	−1.560
25th percentile	−0.674	−0.693
50th percentile	0.000	−0.014
75th percentile	0.674	0.644
95th percentile	1.645	1.653
99th percentile	2.326	2.337

Average V_{it} = cross-sectional average percentage of shares traded

SV_{it} = standardized excess volume measure, computed by standardizing V_{it} by its estimated mean and standard deviation

Z_{SV_t} = statistic generated from SV_{it} by standardizing by the theoretical mean and standard deviation

NT = significance level of Z_{SV_t} based on the standard normal distribution

ED = significance level of Z_{SV_t} based on the emiprical distribution

no evidence for increased trading volume around the SEC filing date, but there is evidence of a significant increase ($p = 0.001$) around the WSJ earnings announcement date.

Taken together, the price and volume results suggest that there was no market reaction in response to the SEC filings for the sample firm-quarters, but there were significant price and volume reactions around the subsequent WSJ earnings announcement dates.

The absence of a market reaction at the SEC filing date may be due to misspecification of the event date. Although a filing becomes public information as soon as it is received at the SEC, public access to the filing is usually delayed one or two days. One proxy for the date that the filing becomes available to the public is the date the filing is received by Disclosure, Inc. (Stober 1986). The tests for a market reaction to the SEC filing were repeated with the SEC interval defined as the three-day period beginning with the receipt of the filing by Disclosure, Inc., which resulted in a sample size of 271 (268 for the volume test). For the Disclosure, Inc., period, the values for Z_{SR_t} (measuring abnormal price activity) and Z_{SV_t} (measuring abnormal volume activity) were -2.060 and -0.542, respectively. As with the values obtained using the SEC receipt date, these are lower than the mean values for the entire test period and again offer no evidence for a market reaction in response to the SEC filing. The market reaction tests were repeated using five-day intervals; the results did not differ from the three-day results. The tests were also repeated on a subset of the sample for which the SEC and WSJ dates were separated by at least eight trading days. Again, the results did not differ from the results for the entire sample. Returns and trading volume for the period in between the SEC and WSJ intervals—SEC_{+3} through WSJ_{-2}—were examined. Neither the returns nor the volume measure were elevated during this interim period.

In summary, there is evidence of the existence of a market reaction to the WSJ earnings announcements for the sample firm-quarters, even when that announcement follows the 10-K or 10-Q filing by as many as eight trading days or more. There is no evidence of the existence of a market reaction, on average, at the SEC filing dates, which were the first public announcements of earnings for the sample firms. There is also no evidence of a market reaction at the Disclosure, Inc., receipt date, nor for the period between the SEC and WSJ intervals.

The existence of a market reaction to the WSJ earnings announcement may indicate the existence of additional price-relevant disclosures at the time of the WSJ announcement. Therefore, the actual WSJ item comprising the earnings announcement was examined for all firm-quarters for which the squared return residual $[(SR_{it})^2]$ for the WSJ interval was greater than four. Thirty-six firm-quarters were in this category, and the appropriate WSJ article was located for all 36. In each case, the earnings announcement was made as part of the "Digest of Corporate Earnings Reports." No cases included any supplemental information. Also, because of the sample selection process, none of the sample firm-quarters had any potentially confounding events announced in the WSJ within two weeks of the SEC or WSJ dates.

The procedures described above cannot eliminate concurrent disclosures as the cause of the market reaction at the WSJ date since Wright and Groff (1986) and Thompson et al. (1987) demonstrate that not all price-relevant disclosures are reported in the WSJ. Further evidence confirming the presumption that the market reaction at the WSJ date is related to the financial data in the earnings announcement can be obtained by verifying a directional relationship between unexpected earnings and abnormal returns for the WSJ interval. Because analysts' forecasts of earnings were unavailable for most of the sample firms, quarterly earnings from the prior year were used as a proxy for expected earnings. Earnings for the sample quarter and for the corresponding quarter in the preceding year were found for 275 of the sample firm-quarters. Of these, 115 represented earnings increases and 160 represented earnings decreases. The following statistic is used to determine whether the positive and negative unexpected earnings firms have different mean abnormal returns for a given interval:

$$Z_{DRt} \equiv \frac{\frac{1}{N_p}\sum_{j=1}^{N_p} SR_{jt} - \frac{1}{N_n}\sum_{k=1}^{N_n} SR_{kt}}{(1/N_p+1/N_n)^{1/2}}, \tag{6}$$

where j subscripts the 115 positive unexpected earnings firm-quarters; k subscripts the 160 negative unexpected earnings firm-quarters; and N_p and N_n are 115 and 160, respectively. Z_{DRt} has an asymptotic standard normal distribution. The values for Z_{DRt} for the 20 three-day intervals are reported in table 4, panel A. Also reported are the cross-sectional average raw returns for the positive and negative unexpected earnings portfolios. The empirical distribution of Z_{DRt} is summarized in panel B. For the WSJ interval, the return residuals for the positive unexpected earnings firm-quarters are shown to be significantly greater than the return residuals for the negative unexpected earnings firm-quarters using both the standard normal distribution ($p=0.008$) and the empirical distribution ($p=0.003$).[6] The nature of the price reaction at the WSJ date is thus shown to be associated with unexpected earnings. This result is consistent with the interpretation that the significant price and volume reactions to the WSJ earnings announcements for the sample firm-quarters can be attributed to the earnings announcements themselves.

III. Summary and Conclusions

The evidence presented in this article shows significant market reactions to WSJ earnings announcements, but not to the filing of 10-K and 10-Q reports with the SEC when such filing preceded earnings disclosure in the WSJ. The reaction existed even when at least eight trading days intervened between the initial public release of earnings in the SEC filing and the subsequent announcement in the WSJ. The sign of the price reaction at the WSJ date was shown to be related to the sign of unexpected earnings.

The results are to be interpreted in light of the nature of the firms in the sample: almost all are small. Only one of the sample firms would be considered large by the criterion used in Atiase (1985), which is a market value of $400 million or greater. Failure to find a market reaction at the SEC filing date may not be duplicated if the sample were drawn from firms followed closely by analysts, e.g., *Fortune 500* firms.

The occurrence of cases in which the 10-K or 10-Q filing precedes the corresponding WSJ earnings announcement is infrequent and the nature of the market reaction in these specific cases is of minor importance. However, two general implications of the results are of greater importance. First, the results suggest that the method of disclosing and cost of acquiring accounting earnings impact whether the information embodied

---

[6] The lack of a significant market reaction overall for the interval SEC_{-6} through SEC_{-4} (see tables 2 and 3) suggests that the reversed relationship between returns and unexpected earnings for this interval is a result of random sampling error.

Table 4

Tests for a Relationship between Security Returns and Unexpected Earnings around the SEC Filings and the WSJ Announcement

Panel A. Test Results Based on 275 Sample Firm-Quarters:

Three-Day Interval	*Raw Returns*			PR($Z_{DRt} \leq$ expected value)	
	Pos UE (n=115)	*Neg UE* (n=160)	Z_{DRt}	*NT*	*ED*
SEC_{-27}–SEC_{-25}	0.009	0.002	−0.809	0.791	0.815
SEC_{-24}–SEC_{-22}	0.010	0.005	0.467	0.320	0.290
SEC_{-21}–SEC_{-19}	0.002	−0.001	0.462	0.322	0.292
SEC_{-18}–SEC_{-16}	0.001	0.001	−1.264	0.897	0.911
SEC_{-15}–SEC_{-13}	0.001	0.003	−0.326	0.628	0.642
SEC_{-12}–SEC_{-10}	0.003	−0.005	0.138	0.445	0.434
SEC_{-9}–SEC_{-7}	−0.004	−0.002	−0.328	0.628	0.643
SEC_{-6}–SEC_{-4}	−0.006	0.008	−2.242	0.988	0.990
SEC_{-3}–SEC_{-1}	−0.006	0.000	−0.684	0.753	0.772
SEC_{0}–SEC_{+2}	0.001	0.000	0.159	0.437	0.424
WSJ_{-1}–WSJ_{+1}	0.014	−0.003	2.412	0.008	0.003
WSJ_{+2}–WSJ_{+4}	0.011	0.002	−0.225	0.589	0.602
WSJ_{+5}–WSJ_{+7}	0.006	0.006	0.177	0.430	0.417
WSJ_{+8}–WSJ_{+10}	0.012	0.000	1.054	0.146	0.103
WSJ_{+11}–WSJ_{+13}	0.009	0.003	1.086	0.139	0.096
WSJ_{+14}–WSJ_{+16}	0.002	−0.001	−0.093	0.537	0.546
WSJ_{+17}–WSJ_{+19}	0.002	0.003	−0.628	0.735	0.759
WSJ_{+20}–WSJ_{+22}	−0.001	0.007	−0.568	0.715	0.736
WSJ_{+23}–WSJ_{+25}	0.001	0.004	−0.011	0.504	0.507
WSJ_{+26}–WSJ_{+28}	0.015	0.004	1.224	0.110	0.072

Panel B. Empirical Distribution of Z_{DRt} Based on 999 Simulated Observations:

	Standard Normal	*Empirical*
Mean	0.000	−0.045
Standard Deviation	1.000	0.897
1st percentile	−2.326	−2.328
5th percentile	−1.645	−1.644
25th percentile	−0.674	−0.609
50th percentile	0.000	−0.003
75th percentile	0.674	0.552
95th percentile	1.645	1.349
99th percentile	2.326	1.899

Raw Returns = cross-sectional average raw return for the positive unexpected earnings (Pos UE) and negative unexpected earnings (Neg UE) subsets of the sample

Z_{DRt} = statistic generated by differencing standardized return residuals (SR_{it}) for the Pos UE and Neg UE firm-quarters

NT = significance level of Z_{DRt} based on the standard normal distribution

ED = significance level of Z_{DRt} based on the empirical distribution in those earnings will be reflected in security prices in a timely fashion. Second, the results suggest that, at least in one specific set of circumstances, data disclosed as part of an SEC-mandated filing are not fully reflected in prices until a subsequent media disclosure is made. The extent to which prices reflect data contained in other SEC filings (especially filings made by small firms) remains an area for future research.

References

Atiase, R. 1985. Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research* 23 (Spring): 21–36.

Bamber, L. 1986. The information content of annual earnings releases: A trading volume approach. *Journal of Accounting Research* 24 (Spring): 40–56.

Beaver, W. 1968. The information content of annual earnings announcements. *Journal of Accounting Research* 6 (Supplement): 67–92.

———. 1981. Market efficiency. *The Accounting Review* 56 (January): 23–37.

Chambers, A., and S. Penman. 1984. Timeliness of reporting and the stock price reaction to earnings announcements. *Journal of Accounting Research* 22 (Spring): 21–47.

Foster, T., and D. Vickrey. 1978. The incremental information content of the 10-K. *The Accounting Review* 53 (October): 921–34.

Hughes, J., and W. Ricks. 1987. Associations between forecast errors and excess returns near to earnings announcements. *The Accounting Review* 62 (January): 158–75.

Lobo, G., and A. Mahmoud. 1989. Relationship between differential amounts of prior information and security return variability. *Journal of Accounting Research* 27 (Spring): 116–34.

Morse, D. 1981. Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research* 19 (Autumn): 374–83.

Noreen, E. 1989. *Computer-Intensive Methods for Testing Hypotheses: An Introduction.* New York: Wiley.

Stober, T. 1986. The incremental information content of financial statement disclosures: The case of LIFO inventory liquidations. *Journal of Accounting Research* 24 (Supplement): 138–60.

Thompson, R., C. Olsen, and J. R. Dietrich. 1987. Attributes of news about firms: An analysis of firm-specific news reported in *The Wall Street Journal Index. Journal of Accounting Research* 25 (Autumn): 245–74.

Wilson, G. P. 1987. The incremental information content of the accrual and funds components of earnings after controlling for earnings. *The Accounting Review* 62 (April): 293–322.

Wright, C., and J. Groff. 1986. Uses of indexes and data bases for information release analysis. *The Accounting Review* 61 (January): 91–100.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 56–66

Supplemental Data and the Structure of Thrift Share Prices

Mary E. Barth
Harvard University
William H. Beaver
Stanford University
Christopher H. Stinson
Stanford University

SYNOPSIS: This study examines thrifts' supplemental disclosures with respect to default risk (scheduled items) and interest rate risk (repricing data). It represents an extension of a recent study (Beaver, Eger, Ryan, and Wolfson 1989; hereafter BERW), which also examines the relationship between banks' share prices and supplemental disclosures with respect to default risk (nonperforming loans) and interest rate risk (maturity data). However, there are differences between the research designs and the findings of the studies.

The research design differences between the bank and thrift studies include: (1) the use of per share deflation instead of book value deflation; (2) the use of seemingly unrelated regressions (SUR), as well as a fixed-effects model; and (3) a partitioning of the sample according to availability of supplemental data to assess the impact of nondisclosure on the estimated coefficients.

A sample of 165 publicly traded thrifts was used to regress market values on several variables, including supplemental disclosures. The findings differ from those of the bank study in several respects. First, the coefficient on the supplemental disclosure on interest rate risk is significant in the case of the banks but not in the case of thrifts. Second, the coefficient on the default risk variable is smaller for thrifts than for banks. Third, the following three additional findings extend the bank study. (1) Thrifts that do not disclose the default risk variable appear to be valued at a discount relative to disclosing thrifts. (2) The estimated effects of the incremental explanatory power of the default risk variable is robust with respect to two nonnested estimation methods, a fixed effects model, and a seemingly

Submitted November 1988.
Revisions received May 1989, and February and August 1990.
Accepted February 1990.

unrelated regression. (3) The finding regarding scheduled items is robust with respect to an alternative specification, which assesses differences in estimated coefficients for "good" loans versus "bad" loans.

Key Words: *Share price, Scheduled items, Repricing data, Thrifts.*

Data Availability: *Inquiries regarding data availability should be submitted to the authors together with a brief description of the intended uses of the data.*

A description of the research proceeds in six steps: (1) a discussion of the differences between banks and thrifts, (2) the nature of supplemental data reported in the thrift industry, (3) a model of market value of common equity and supplemental disclosures, (4) the selection of the sample, (5) a discussion of the empirical results, and (6) concluding remarks.

I. Differences Between Thrifts and Banks

Notwithstanding the fact that banks and thrifts share many common features, thrifts possess several distinct characteristics. These distinctions relate to the nature of: (1) the transactions and assets held, (2) the regulatory environment, (3) the disclosure requirements and practices, and (4) the financial strength of the industry.

In terms of transactions, thrifts' asset structure consists of a larger portion of collateralized, residential mortgages than does that of banks. While the collateralized nature of the loan may suggest that default risk is relatively smaller for thrifts, in practice the economic value of the collateral tends to be highly correlated with the earning power of the mortgagee. The average duration of the mortgage loan is also much longer than the average duration of a commercial bank loan.

In terms of the regulatory environment, banks tend to be supervised by federal banking agencies. Thrift regulation has tended to be placed more in the hands of the regional districts of the Federal Home Loan Bank Board (FHLBB). This "home rule" aspect of thrift regulation in part led to the recent policy to move liquidation of insolvent thrifts from the FHLBB to the Federal banking regulators (Howe 1989). This difference in regulatory climate is also reflected in differences in reporting practices, especially the mandatory versus voluntary nature of the disclosures.

In particular, while the default risk disclosure for banks (nonperforming loans) is required, disclosure of the analogous item (scheduled items) for thrift holding companies is not. Interest-rate risk disclosures (repricing data) are required for banks but not for thrifts. The more "relaxed" voluntary nature of thrift disclosure policies may induce systematic pricing differences between nondisclosers and disclosers, an aspect that is explored in the research design.

Not only are there differences in the voluntary nature of disclosure policies but also in the nature of the disclosed data. Banks' *nonperforming loans* include nonaccrual loans and loans restructured because of financial difficulty. In contrast, *scheduled items*, the analogous thrift supplemental disclosure, includes delinquent and substandard loans. These concepts can be considerably different. For example, in 1987 Financial Corporation of America (FCA) reported scheduled items of $1.05 billion but nonperforming loans of $2.9 billion.[1] Moreover, in assessing allowance for loan losses, thrifts may apply a net realizable value (NRV) criterion to specific loans, while banks do not (AICPA 1986). However, even in the case of thrifts, the NRV criterion is applied only to a subset of loans (i.e., loans for which there is a specific allocation to the allowance account). Thus the impairment due to default risk is only partially reflected. Moreover, there is no attempt to adjust the loans to NRV to reflect any impairments due to interest-rate risk (i.e., unexpected changes in interest rates).

Thrifts are in deeper financial difficulty than banks.[2] For firms experiencing financial difficulty, the optionlike characteristics of its common stock are likely to become a more important factor in explaining share price. As a result, supplemental data that reflect asset impairment may also be reflecting the value of an intangible asset (i.e., the option to "put" an insolvent institution to a federal insurance corporation or uninsured liability holders). The value of the intangible put is expected to vary inversely with supplemental data reflecting asset impairment, potentially driving the coefficient estimate on the supplemental data toward zero.

II. Financial Reporting of Supplemental Data

The thrift industry can be characterized as being exposed to default risk and interest-rate risk. For reasons discussed at length in BERW, the allowance for loan losses may not fully reflect the economic impairment of loans. As a result the market value of the loan can fall below the reported book value. To the extent that thrifts follow a net realizable value criterion, this economic impairment may be partially reflected in the allowance for loan losses.[3] However, as indicated earlier, the economic impairment generally may not be fully reflected. For example, FASB Statement No. 5 requires the loss to be "probable" before it can be recorded in the allowance account. Moreover, market values of the loans will vary when the discount rate unexpectedly changes (i.e., interest-rate risk) and are not reflected in the allowance account or any other account.

Interest-rate risk occurs when the duration of interest-bearing assets is not equal to that of interest-bearing liabilities. Historically, thrifts have been viewed as a classic case of interest-rate mismatch because they are long on the asset side and short on the liability side. In particular, unexpected increases (decreases) in the interest rate decrease (increase) the discounted present value of the loan portfolio. Generally Accepted Accounting Principles (GAAP) do not require that the economic impairment in the loan be recorded.

---

[1] Nonperforming loans are generally not disclosed by thrifts and hence were not available for study.

[2] From 1980 through 1985, 581 thrifts failed, representing approximately 15 percent of thrifts existing in 1980. During the same time period, 309 banks failed representing two percent of the banks existing as of 1980 (Barth et al. 1986). As of year-end 1987, the Federal Deposit Insurance Corporation (FDIC) had a surplus of approximately $18 billion, while FSLIC was estimated to have a deficit of $12.5 billion (Hilder 1988). In 1987, one-third of the approximately 3,000 thrifts had negative earnings. Banks also reported losses in 1987. However, the losses were primarily due to losses recorded on loans to developing countries by the largest 25 banks. The rest of the industry operated at a profit.

[3] The definition of scheduled items has varied over time, but major categories include "slow" consumer credit, "slow" conventional mortgage loans, 20 percent of insured or guaranteed loans, and real estate owned (REO) as a result of the foreclosure process. In 1986, the definition was changed to include substandard assets and to exclude "slow" mortgage loans, except for residential loans secured by fewer than five housing units. Effective September 30, 1987, the concept of scheduled items was eliminated in favor of a related concept, classified assets.

The recent use of adjustable-rate mortgages (ARMs) and financial futures markets permits lenders to reduce interest-rate risk. Supplemental disclosures often report interest-rate sensitivity in a repricing schedule, which lists interest-rate-sensitive assets and liabilities by major categories and by repricing periods. For adjustable-rate mortgages, the repricing period is the earliest time at which the interest rate can be adjusted (typically less than one year). For fixed-rate mortgages, the repricing period is based on maturity as adjusted for expected prepayment of principal.

ARMs are restricted in terms of frequency of revision, the maximum the interest rate can change at any repricing point, and the maximum interest rate attainable under the mortgage. Given such limitations, a 30-year adjustable-rate mortgage with repricing every six months is not of equivalent interest-rate risk to a six-month certificate of deposit, although a repricing schedule would assign them to the same repricing period. Despite the complexity of the concept of interest-rate exposure, we follow prior research by defining interest-rate exposure in terms of the loans subject to repricing beyond one year.

Unrecorded economic impairment due to unexpected increases in default risk and/or in interest rates can cause the market value of thrifts' common equity to be less than the book value recorded under GAAP. The primary purpose of this study is to examine the extent to which supplemental disclosures, such as scheduled items and repricing data, provide incremental explanatory power with respect to the differences between the market values and book values of thrifts. The next section explicitly models these relationships.

III. Market Value of Common Equity and Supplemental Disclosures

This section relates the market value of thrifts' common equity to supplemental disclosures. The model presented here is an extension of the model developed in BERW.

$$CE^m = CE^b + L^m - L^b + ONA^m - ONA^b, \tag{1a}$$

or equivalently,

$$CE^m - CE^b = L^m - L^b + ONA^m - ONA^b, \tag{1b}$$

where CE denotes common shareholders' equity; L denotes loan assets; ONA denotes all non-loan assets, net of all liabilities and preferred stock; and superscripts m and b denote market value and book value, respectively.

The valuation difference between the market value and book value of common equity ($CE^m - CE^b$) is (tautologically) equal to the valuation differences in the loan portfolio ($L^m - L^b$) and in other (net) assets ($ONA^m - ONA^b$). Note that L^m is not observable to the researcher. The discussion in the previous section suggests that the valuation error in loans (i.e., $L^m - L^b$) should be related to the magnitude of scheduled items, as well as to the repricing structure of the loans. Specifically, assume:

$$L^m - L^b = \gamma_0 + \beta_1 ALL + \beta_2 SI + \beta_3 REP + \epsilon, \tag{1c}$$

where ALL, SI, and REP denote the book values of allowance for loan losses, scheduled items, and loans with repricing periods greater than one year. Further assume that:

$$ONA^m - ONA^b = \delta_0 + \mu CE^b + \nu, \tag{1d}$$

where ν is independent of *ALL*, *SI*, and *REP*, and has a mean of zero. μ reflects the valuation difference per dollar of common equity. Substituting equations (1c) and (1d) into equation (1a) yields:

$$CE^m = \beta_0 + \beta_1 ALL + \beta_2 SI + \beta_3 REP + \beta_4 CE^b + u, \tag{1e}$$

where

$$\beta_0 = \gamma_0 + \delta_0,$$
$$\beta_4 = 1 + \mu, \text{ and}$$
$$u = \epsilon + \nu.$$

ALL is included in the specification of equations (1c) and (1e) to capture the portion of $L^m - L^b$ that can be explained simply as a linear function of *ALL* (i.e., *ALL* times some proportionality factor). This is important because the purpose of the study is to examine the incremental explanatory power of supplemental data beyond that provided by the financial statement account, allowance for loan losses. Plausible arguments could be offered for β_1 being negative, zero, or positive.

The arguments presented earlier can be interpreted as predicting that $\beta_2 < 0$ (i.e., the higher SI, the lower $L^m - L^b$). As discussed earlier, if the value of deposit insurance increases inversely with the value of impaired assets, estimates of β_2 could be driven toward zero. In principle, the value of the put option is nonlinear and related to the "exercise price" (i.e., total deposits and liabilities). In this respect, equation (1c), which posits a linear relation between the valuation error and the supplemental data, is to be viewed as an approximation. The repricing variable (*REP*) is intended to capture valuation errors induced by unanticipated changes in interest rates since the inception of the mortgages. If market interest rates have increased unexpectedly since the origination or acquisition of the loans (as implied by the industry experts' estimates of economic losses due to interest-rate changes cited earlier), β_3 would be expected to be negative.

$\beta_4 = 1 + \mu$, where μ reflects the systematic component of the difference between the market value and book value of net assets other than loans. There is no prediction regarding the sign of μ.

A nonzero intercept in either equation (1c) or (1d) would produce an intercept (β_0) in equation (1e). The intercept terms represent the mean effect of the respective valuation differences that are not captured in the proportional relationships to the right-hand side variables in equations (1c) and (1d).

IV. Sample Selection

A comprehensive list of publicly-traded thrifts is not readily available. As a result, several sources were consulted.[4] These sources yielded 492 thrifts, including many that are not publicly traded. Since the market price of common stock was a data requirement, the ISL Daily Stock Price Record was examined to determine the availability of stock price quotations. This criterion precludes publicly-owned stocks with limited trading and yielded a sample of 203 thrifts.

---

[4] The sources included: a list of thrift holding companies that file with the SEC (77 firms), a list of thrifts that are required to file reports (for Securities Acts' purposes) with the FHLBB as of April 1986 (324 firms), Moody's, the *American Banker* list of top 100 thrifts, COMPUSTAT tapes, the sample used by Brickley and James (1986), the NAARS database, Peat Marwick's publication *Principles and Presentation: Thrifts*, Paine Webber's list of thrift stocks, and firms in the Thrift Index compiled by Keefe, Bruyette, & Woods.

Table 1
Means of Selected Variables
1980–1987

	Years							
	1980	1981	1982	1983	1984	1985	1986	1987
CE^m/CE^b	0.66	0.50	0.85	0.82	0.70	0.88	1.00	0.75
NI/CE^b	0.08	−0.09	−0.13	0.08	−0.01	0.10	0.11	0.02
SI/CE^b	0.12	0.29	0.64	0.39	0.59	0.58	0.73	1.26
ALL/CE^b	0.01	0.02	0.03	0.03	0.06	0.07	0.12	0.13
REP/CE^b	N.A.	N.A.	4.01	16.84	30.62	16.79	11.32	12.32
CE^b/TA	0.06	0.04	0.05	0.05	0.05	0.05	0.05	0.05
NOBS(SI)	23	30	39	77	121	147	146	96
NOBS(REP)	0	0	1	8	28	84	107	81
NOBS	35	36	44	84	132	155	155	137

CE^m = Market value of Common Equity
CE^b = Book value of Common Equity
SI = Scheduled Items
ALL = Allowance for Loan Losses
REP = Interest-rate sensitive assets with repricing periods greater than one year
TA = Book value of Total Assets
NOBS = Number of observations for scheduled items, repricing variable, and total, respectively
N.A. = Not computable
If $CE^b < 0$, observation was omitted in computation of mean.

Requests for annual reports were sent to the 203 thrifts. Of the 176 responses obtained, 165 supplied annual reports.[5] Of these, 24 are New York Stock Exchange (NYSE), eight are American Stock Exchange (AMEX), and 133 are Over-The-Counter (OTC) firms. The data for the financial reporting variables were obtained from the annual reports. When the variables were not disclosed in the annual report, the SEC 10-K report was examined in an attempt to obtain the information. Price data were obtained from the ISL Daily Stock Price Record.

V. Empirical Analysis

The analysis of the empirical results proceeds in three parts: (1) estimation of supplemental data coefficients in the basic model, (2) fixed-effects estimation, and (3) estimation of the effects of nondisclosure.

Estimation of Supplemental Data Coefficients in the Basic Model

Table 1 reports the number of observations and the mean values of selected financial statement variables for the years 1980–1987. An analysis of mean values is discussed in an earlier version of this research (Barth et al. 1989). The earlier analysis also

---

[5] Eleven firms responded but did not supply annual reports. They claimed they were not "publicly traded" because of concentration of ownership by an individual, corporation, or investment group.

Table 2
Regression Coefficients for 1983–1987 on Sample Data for which Repricing Data are Available*

	Years					*Fixed Effects*
	1983	*1984*	*1985*	*1986*	*1987*	
Constant	6.10 (0.46)	2.01 (0.86)	1.39 (0.67)	9.30 (4.78)	11.16 (5.97)	N.R.
ALL/SHR	8.59 (0.55)	3.63 (1.28)	0.95 (1.27)	0.49 (1.80)	0.57 (1.14)	0.44 (1.90)
SI/SHR	−0.64 (−0.40)	−0.37 (−1.09)	−0.04 (−0.43)	−0.14 (−1.81)	−0.15 (−1.99)	−0.04 (−0.57)
REP/SHR	−0.02 (−0.36)	0.007 (0.86)	−0.002 (−0.32)	−0.01 (−1.26)	−0.02 (−2.56)	0.004 (1.01)
CE^b/SHR	0.70 (1.44)	0.45 (3.36)	0.72 (8.95)	0.53 (6.51)	0.33 (4.44)	0.47 (7.06)
NOBS	7	27	83	103	59	285
Adj. R^2	−0.26	0.26	0.53	0.36	0.34	0.68

* Regression coefficients obtained from the following regression (*t*-statistics in parentheses):

$$CE^m/SHR = \alpha + \beta_1 ALL/SHR + \beta_2 SI/SHR + \beta_3 REP/SHR + \beta_4 CE^b/SHR + u.$$

N.R. = Not reported

demonstrated that estimation in per-share form was a more appropriate adjustment for heteroskedasticity than book-value deflation. As a result, the estimating equation is:

$$\frac{CE^m}{SHR} = \beta_0 + \beta_1 \frac{ALL}{SHR} + \beta_2 \frac{SI}{SHR} + \beta_3 \frac{REP}{SHR} + \beta_4 \frac{CE^b}{SHR} + u, \tag{1f}$$

where:

SHR = number of common shares outstanding.

β_1 through β_4 are expected to have the same sign as in equation (1e).

Table 2 reports the results of estimating equation (1f) for 1983 through 1987. There are insufficient degrees of freedom in 1980 through 1982 to estimate coefficients, largely because of the lack of data availability of the repricing variable. Even in 1983 there are only two degrees of freedom. The coefficient on CE^b is positive and significant in 1984 through 1987, while *ALL* and *REP* are usually insignificant. The sign of the *SI* coefficient is negative as predicted in every year and is significant in 1986 and 1987.[6]

---

[6] Without relying on asymptotic theory, taking the *t*-values for a given year at face value assumes that the residuals are normally and independently distributed. Hansen (1982) discusses the asymptotic properties of ordinary least squares regressions. Cross-sectional correlation is often raised as a concern. If each year's data were independent of previous years, a significance test might be constructed based on the time-series standard deviation of the regression coefficients. However, as reported later, the evidence indicates that the residuals from different year's regressions are correlated. In assessing the significance of a coefficient, significance at the conventional five percent level will be used. The tests are one-sided. However, given the results reported, the decision to choose a one-tailed or two-tailed test does not affect the basic conclusions.

Table 3

Regression Coefficients for 1980–1987 for Sample Data where Availability of Repricing Data is Not Required*

	Years								Fixed Effects
	1980	1981	1982	1983	1984	1985	1986	1987	
Constant	8.62	5.14	8.21	6.78	5.44	6.66	9.62	9.71	N.R.
	(3.82)	(2.51)	(3.24)	(4.05)	(3.90)	(4.71)	(5.92)	(6.66)	
ALL/SHR	0.62	2.25	0.99	0.53	0.10	−0.20	0.43	−0.02	0.03
	(0.58)	(2.37)	(1.05)	(0.60)	(0.18)	(−0.43)	(1.69)	(−0.05)	(0.18)
SI/SHR	−1.02	−0.34	−0.28	−0.11	−0.04	0.01	−0.22	−0.15	−0.14
	(−2.75)	(−1.46)	(1.32)	(−0.97)	(−0.40)	(0.07)	(−3.40)	(−2.76)	(−4.18)
CE^b/SHR	0.29	0.24	0.31	0.40	0.34	0.48	0.48	0.26	0.47
	(3.35)	(2.87)	(2.82)	(5.43)	(6.34)	(9.12)	(7.54)	(4.97)	(11.95)
$NOBS$	21	29	38	75	119	147	145	95	669
Adj. R^2	0.47	0.33	0.22	0.29	0.25	0.38	0.35	0.34	0.65

* Regression coefficients obtained from the following regression (*t*-statistics in parentheses):

$$CE^m/SHR = \alpha + \beta_1 ALL/SHR + \beta_2 SI/SHR + \beta_3 CE^b/SHR + u.$$

N.R. = Not Reported

Losing so many observations because of the nonavailability of one variable is a high price to pay. The coefficients are reestimated without the repricing variable and the results are reported in table 3. CE^b is positive and significant in all years. *ALL* is generally positive but insignificant. *SI* is negative except for 1985 and is significant in 1980, 1986, and 1987.[7]

Fixed-Effects Estimation

Each year's results are unlikely to represent a drawing independent of the previous year's. The variables are known to be highly correlated over time. Diagnostics on the behavior of the residual terms indicate that adjacent years' residuals are positively correlated. The correlations range from 0.35 to 0.72 and are significantly different from zero. This suggests the omission of some firm-specific factor that explains CE^m and persists from year-to-year.

A fixed-effects model is used to alleviate this problem. The fixed-effects model is a special case of Generalized Least Squares estimation, where a particular structure is placed on the behavior of the residuals. A fixed-effects model is discussed more fully in BERW and is essentially equivalent to a pooled times-series, cross-sectional estimation with constant slope coefficients, but with time-specific and firm-specific intercept terms. In this spirit, equation (1f) can be expressed (in per-share terms) as:

---

[7] As discussed in section III an intercept in either equation (1c) or (1d) would produce an intercept in equation (1e). However, if the intercept in equation (1e) were nonzero, equation (1f) could be reestimated with an additional explanatory variable equal to $1/SHR$ and the resulting intercept would be expected to be zero. When $1/SHR$ was added to the estimating equation, the coefficient on $1/SHR$ was insignificant and the intercept term remained significant. Measurement error in the explanatory variables is a plausible explanation for the positive intercept. Another possibility is that optionlike characteristics of thrifts' common stocks induce a nonlinear relationship between market value and the explanatory variables.

$$CE^{m}_{it} = \Sigma\alpha_i + \Sigma\alpha_t + \beta_1 ALL_{it} + \beta_2 SI_{it} + \beta_3 REP_{it} + \beta_4 CE^{b}_{it} + u_{it}. \tag{1g}$$

The estimated slope coefficients from the fixed-effects model including the repricing variable are reported in the last column of table 2. CE^b and *ALL* are positive and significant. Neither *SI* nor *REP* is significant. However, requiring the presence of the repricing variable reduces the degrees of freedom and may impose some form of self-selection bias. The coefficients are reestimated without the repricing variable and the results are reported in the last column of table 3. CE^b is positive and significant, *ALL* is positive but insignificant, while *SI* is negative and significant.[8] Taken at face value, a coefficient of −0.14 on *SI* implies that investors, in assessing stock prices, act as if they apply a 14 percent discount to mortgages that are included in scheduled items. The discount is smaller than the 63 percent found for banks' nonperforming loans in BERW.[9]

SI moves from insignificance to significance when *REP* is deleted from the estimating equation. There are at least three possible explanations: (1) The increased number of observations provides a more efficient estimate. (2) *SI* is no longer forced to compete with *REP*, which implies that collinearity might be explaining the insignificant result. (3) Some form of self-selection is present such that those firms that choose to report a repricing variable are those for which *SI* is incrementally less informative. To further explore this issue, the coefficients are estimated without the repricing variable included but on the same set of firms as reported in table 2 (i.e., firms for which *REP* is available). The results are essentially the same as those reported in table 2. This finding implies that the insignificant result for the *SI* coefficient is not due to collinearity between *SI* and *REP*.

Consistent with the self-selection conjecture, the coefficient on *SI* for the larger sample (table 3) is almost four times the −0.04 reported for the smaller sample (table 2). Hence, the significance of the *SI* variable is due in part to an increase in the coefficient as well as a reduction in the standard error because of more observations.

Diagnostics on the residuals indicate that the correlation of adjacent years' residuals is −0.03 and is not significantly different from zero. This finding suggests that the fixed-effects model is successful in removing the source of correlation in adjacent years' residuals.[10]

---

[8] White's (1980) test for heteroskedasticity is conducted. The test statistic compares a heteroskedastic-consistent estimate of the covariance matrix with the "usual" estimate of the covariance matrix, computed under the assumption of homoskedasticity. A function of the difference is shown to be distributed as χ^2. The White test statistic was computed for the estimating equation used in table 3. With three explanatory variables, the White test statistic is distributed χ_2 with nine degrees of freedom, and the critical value at the 0.05 level of significance is 16.9. The White statistic for the fixed-effects model reported in table 3 is a χ^2 value of 36.41, which exceeds the critical (.05) value of 16.9. White estimates of the standard errors imply *t*-values of 0.24, −8.18, and 29.84 for *ALL*, *SI*, and CE^b, respectively. The *t*-value for *ALL* is basically unchanged, while the *t*-values for SI and CE^b are at least double those reported in table 3.

[9] Caution must be exercised when drawing inferences about undeflated variables from coefficients derived from deflated data. The caveat applies doubly here because different deflators are used in both studies. In making such comparisons, we assume the deflation makes a heteroskedasticity adjustment so that unbiased but more efficient estimates of the coefficient for the undeflated variables are obtained. These issues are discussed in greater length in Lev and Sunder (1979) and Christie (1987).

[10] The coefficients were also estimated using seemingly unrelated regressions (Schipper and Thompson 1983). The results regarding the significance of scheduled items is robust with respect to fixed-effects and SUR estimation methods. It would be interesting to compare the two formulations. However, neither is nested within the other, and the comparison of nonnested alternative specifications is difficult and beyond the scope of this study.

Estimation of Effects of Nondisclosure

An expanded version of equation (1g) was considered in order to estimate the effects of nondisclosure on the estimated coefficients. The expanded version is of the form (in per-share terms):

$$CE^m_{it}=\Sigma\alpha_i+\Sigma\alpha_t+\omega_1 DSI_{it}+\omega_2 DREP_{it}+\beta_1 ALL_{it}+\beta_2 SSI_{it}+\beta_3 SREP_{it}+\beta_4 CE^b_{it}+u_{it},$$

where:

$DSI=1$ if scheduled items are not disclosed, 0 otherwise;
$DREP=1$ if repricing data are not disclosed, 0 otherwise;
$SSI=$ scheduled items if disclosed, 0 otherwise; and
$SREP=$ repricing data if disclosed, 0 otherwise.

ω_1 and ω_2 reflect the mean difference in the dependent variable of the disclosing and nondisclosing firms, conditional upon the values of the other explanatory variables. These coefficients reflect at least two effects: the inferred values of the nonreported data and a penalty (reward) attached to nondisclosure (disclosure). The expanded form of the estimating equation is expected to produce more efficient estimates because it captures the effects of nondisclosure and because there is no loss of observations due to lack of data availability of either *SI* or *REP*. The coefficients and respective *t*-values (in parentheses) for the estimated equation are:

$$\begin{array}{l}CE^m=\Sigma\alpha_i+\Sigma\alpha_t-1.61DSI+0.22DREP-0.08ALL-0.114SSI+0.002SREP+0.45CE^b+u.\\ \qquad\qquad\quad(-2.01)\qquad(.25)\qquad(-0.56)\quad(-3.59)\qquad(.75)\qquad(12.17)\end{array}$$

The results are essentially the same as those in table 3. *SSI* and CE^b are significant while *ALL* is not. ω_1 is -1.61, indicating that market price per share is lower for nondisclosing firms. This is consistent with an environment where "good news" is more likely to be disclosed than "bad news" and/or where nondisclosers are penalized for not reporting relevant information. Similar to earlier results, neither coefficient of the repricing variables is significant.[11]

---

[11] Another extension was made at the suggestion of one of the anonymous referees of an earlier version of this paper. The estimating equation is (expressed in per-share terms):

$$CE^m_{it}=\Sigma\alpha_i+\Sigma\alpha_t+\beta_1 TL^b_{it}+\beta_2 SI_{it}+\beta_3 ONA^b_{it}+u_{it}$$

where:

$TL^b=$ total loans (net of allowance for loan losses)

$$ONA^b=CE^b-TL^b.$$

The estimated coefficients are (*t*-values in parentheses):

$$\begin{array}{l}CE^m_{it}=\Sigma\alpha_i+\Sigma\alpha_t+0.475TL^b_{it}-0.306SI_{it}+0.470ONA^b_{it}+u_{it}.\\ \qquad\qquad\qquad\quad(6.82)\qquad(-3.41)\qquad\quad(6.55)\end{array}$$

β_2 is negative and significant, which implies that a lower coefficient is applied to "bad loans" as measured by *SI* than the rest of the loan portfolio, even after deducting the allowance for loan losses. Total loan data were not available for all of the sample firms, and a direct comparison with the results in table 3 is not possible. However, the same fixed-effects estimating equation applied to the observations reported in table 3 was estimated for the observations that were used in this "good loan/bad loan" analysis. The coefficients were 0.520, −0.314, and 0.527, for ALL^b, *SI*, and CE^b, respectively.

VI. Concluding Remarks

This study examines the incremental explanatory power of supplemental data reported by thrifts with respect to scheduled items and repricing data. In estimation techniques with more properly behaved residuals, scheduled items exhibits a significant, negative relationship to the market value of common equity. This finding is robust with respect to a fixed-effects and a SUR specification.

In an extension of the basic research design, nondisclosure of scheduled items is associated with a lower market price per share, conditional upon the other explanatory variables. The finding is consistent with investors drawing inferences about the magnitude of the nonreported data and/or with investors assessing a "penalty" for nondisclosure.

As in the earlier bank study (BERW), the coefficient for supplemental disclosures related to default risk is significant, although the coefficient for thrifts is much lower. This finding is consistent with three explanations: the default risk of collateralized loans is lower; scheduled items measure default risk with greater error; and the value of the insurance put option is relatively greater. Further, while supplemental disclosure regarding interest-rate risk was significant for banks, it was not for thrifts.

Insignificant results are typically difficult to interpret. One explanation is that grouping by earliest date of repricing is too crude a measure of interest-rate risk. Repricing periods ignore the constraints on repricing that may exist (differentially) on assets and liabilities with the same repricing period. Another explanation is that unexpected interest-rate changes were not significant in the 1984–1987 period when the repricing data were reported. Another, less plausible, explanation is that thrifts have no differential interest-rate risk, however measured.

References

AICPA (American Institute of Certified Public Accountants). 1986. *Audit and Accounting Guide: Savings and Loan Associations.* New York: AICPA.

Barth, J., R. Brumbaugh, Jr., and D. Sauerhaft. 1986. *Failure Costs of Government-Regulated Financial Firms: The Case of Thrift Institutions.* Washington, D.C.: Federal Home Loan Bank Board.

Barth, M., W. Beaver, and C. Stinson. 1989. Supplemental data and the structure of thrift share prices. Working paper, Stanford University.

Beaver, W., C. Eger, S. Ryan, and M. Wolfson. 1989. Financial reporting, supplemental disclosures, and the structure of bank prices. *Journal of Accounting Research* 27 (Autumn): 157–78.

Brickley, J., and C. James. 1986. Access to deposit insurance, insolvency rules, and the stock returns of financial institutions. *Journal of Financial Economics* 16 (July): 345–71.

Christie, A. 1987. On cross-sectional analysis in accounting research. *Journal of Accounting & Economics* 9 (December): 231–58.

Hansen, L. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50 (December): 1029–54.

Hilder, D. 1988. Shaky coverage: Thrift-industry crisis may force overhaul of deposit insurance. *The Wall Street Journal.* 27 July, 1.

Howe, K. 1989. How the thrift industry crashed. *San Francisco Chronicle.* 27 February, C1.

Lev, B., and S. Sunder. 1979. Methodological issues in the use of financial ratios. *Journal of Accounting & Economics*: 1 (December): 187–210.

Peat Marwick & Main. 1982–84. *Principles and presentations; Savings and loan.* New York: Peat Marwick & Main.

Schipper, K., and R. Thompson. 1983. The impact of merger-related regulations on the shareholders of acquiring firms. *Journal of Accounting Research* 21 (Spring): 184–221.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (May): 817–38.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 67–79

A Perspective on Accounting and Stock Prices

Michael J. Brennan
University of California, Los Angeles

COMMON sense tells us that accounting information is relevant for valuing stocks. Financial analysts pay large sums for such information, analyze it carefully, and make forecasts about future accounting information (earnings) in justifying their buy and sell recommendations; we also read in the newspaper that major earnings surprises are accompanied by significant changes in stock prices. The scientific study of the relation between stock prices and accounting information commences with the classic paper by Ball and Brown (1968). This study, conceived in the heyday of the efficient markets hypothesis, found that accounting information, more precisely accounting earnings, was part of the information used in forming prices, although the data plots strongly suggested that the market anticipated the bulk of the information in the annual accounting numbers. At the risk of oversimplification, the research that has flowed from this early article can be summarized under two headings: (1) The research on the effects of information, primarily new accounting information, on market prices, trading volume, analysts' forecasts, and other capital market phenomena. The articles in the current Forum by Swaminathan and Stice fall under this heading. (2) The research relating share prices to currently available accounting information. The paper by Barth, Beaver, and Stinson falls into this category.

These two types of study, which for convenience I refer to as market reaction and valuation studies, respectively, are not entirely distinct. Insofar as it is concerned with price effects, the former type generally focuses on the relation between new accounting information and short-term changes in stock prices. Valuation studies, on the other hand, are ostensibly concerned with the relation between the level of stock prices and accounting variables, generally earnings, but in order to mitigate problems arising from omitted variables, valuation studies frequently use first differences and also consider the relation between changes in stock prices and changes in earnings. However, the distinction between the two types of study may be maintained by observing that market reaction studies generally analyze security returns over much shorter time intervals.

The author is grateful to W. Beaver and J. Ohlson for helpful comments on an earlier draft of this paper. However, he accepts responsibility for remaining errors and misunderstanding.

Commissioned by the Editor.
First draft, February 1989.
Revisions received July and August 1990.

Following the broad topics of the Forum, I discuss in the following section a number of studies that relate to the market reaction to the announcement of new accounting information. Section II is concerned with studies that deal with the role of accounting numbers in valuation, and in it I argue that accounting earnings do not correspond well to their theoretical counterpart. Section III is a discussion of some models that have yet to be empirically implemented in which accounting earnings are important in valuation.

I. Market Reactions to New Accounting Information

Under the efficient markets hypothesis, a natural metric for assessing the information content of accounting earnings announcements is the fraction of the annual variance of stock returns that is associated with these announcements.[1] In a classic early study, Beaver (1968) found that the variance of residual returns was about 67 percent above normal in the week of an annual earnings announcement and about ten percent above normal in the preceding and following weeks. Taking account of the three weeks surrounding the earnings announcement, this suggests that the announcement accounts on average for about one week's normal residual return, or about two percent of the annual variance of the residual return. Even taking account of quarterly earnings announcements, it seems unlikely that earnings announcements in aggregate could account for more than about five to six percent of the annual residual variance. This is consistent with the picture in Ball and Brown (1968), which reveals that most of the price action precedes the announcement. However, the results of Atiase (1985) suggest that results based on "averages" will conceal size-related cross-sectional variation.

Not only is a small part of the annual residual variance associated with the earnings announcement period, but cross-sectional regressions of announcement period returns on measures of unexpected earnings have low explanatory power[2] so that earnings announcements explain little of the return over the announcement period. This suggests that accounting earnings are not relevant for valuation, or that their announcement is swamped by other information that is relevant for valuation, or that the market treats deviations of reported earnings from prior forecasts as transient deviations.

On the other hand, the efficient markets assumption that stock prices change only because of the arrival of new information is becoming increasingly difficult to sustain. Roll (1988) shows that there is considerable variance in daily returns that cannot plausibly be attributed to the arrival of new (public) information.[3] His results suggest that new, firm-specific information accounts on average for an increase in the daily variance of only about 12 percent. The implied background noise in security returns, which is not related to news, is consistent with the rational expectations models of Grundy and McNicholls (1989) and Wiang (1989), which allow for both informed and uninformed rational traders and noise traders. It is the competitive noisy rational expectations equilibrium models (NREE) of Hellwig (1980) and others that underlie these papers, and the market microstructure models of Kyle (1985), Glosten and Milgrom (1985), Diamond and Verrecchia (1987), and others that currently appear to offer the best prospects for developing our understanding of the means by which new information comes to be reflected in securities prices; the periodic announcement of standardized accounting information provides an almost perfect set of data on which to study the effect of new public information on security prices and trading.

---

[1] This abstracts from the problem that other information is likely to be released at the same time as the earnings announcement.

[2] See Lev (1988) for a critical survey.

[3] See also French and Roll (1986).

The Forum article by Swaminathan in this issue uses a noisy rational expectations model (NREE) of the Hellwig type to arrive at testable implications about the relation between the precision of earnings announcements on the one hand, and the variability of stock prices and the divergence of posterior beliefs on the other. These hypotheses are tested using segmented earnings reports as a proxy for more precise information. The findings support the hypotheses that more precise information leads to greater variance in stock prices and more agreement across analysts. These results are interesting, and the attempt to derive testable hypotheses from the formal model is laudable. Nevertheless, it must be pointed out that the formal model treats the earnings announcement as a *private* signal and allows for only a single round of trading. As the author makes clear, an increase in the precision of *iid* private signals given common priors can increase the dispersion of posterior beliefs if the stock price is a poor aggregator of private information because of supply noise. Consequently, the model itself does not yield unambiguous predictions about the sign of the effect of an increase in signal precision on either price variability or the divergence of posterior beliefs; instead, the hypotheses have to be derived by imposing *ad hoc* restrictions on the model parameters to yield predictions that depend not upon the model, though they are not inconsistent with it, but upon the author's prior (reasonable) conjectures. It might be more appropriate to cast the analysis in terms of a NREE with multiple rounds of trading and to analyze explicitly the effect of the precision of a noisy public signal. Such an analysis is possible within the framework of Grundy and McNicholls (1989), who characterize the volume of trading on release of a public signal. In their model, post-signal beliefs are congruent as the second round of trading reveals the earlier private information.

The paper by Stice has a more explicit efficient markets orientation. He demonstrates quite conclusively that, for a sample of small firms whose earnings are announced late in *The Wall Street Journal* (*WSJ*) relative to their 10-K filing, the market reacts not to the prior 10-K filing but to the *WSJ* announcement. The efficient markets apologist might be tempted to ask why it is that some *WSJ* announcements are delayed. Is it simply a matter of space in the newspaper, or does the *WSJ* delay signal some unobserved obstruction to the timely dissemination of information? Assuming that there is no significance to the *WSJ* delay, the results are in nice accord with recent developments in the theory of trading in securities markets that take account of information asymmetries between market participants.[4] These theories would predict that uninformed "liquidity" traders who have the ability to postpone their trading would tend not to trade immediately before an important news announcement since this is the time when their information disadvantage is greatest. On the other hand, they would tend to increase their trading in the period immediately after the announcement when their information disadvantage is likely to be least. At the same time, traders who have superior information will also tend to time their trading to coincide with that of the uninformed traders, for when the market is deepest their price impact is likely to be least.[5]

---

[4] See Kyle (1985), Admati and Pfleiderer (1989), and Foster and Viswanathan (1990).

[5] Offsetting this to some extent is the decay in their informational advantage due to the impending announcement.

These tendencies mean that one would expect a reduction in trading before a public announcement and an increase in trading following the announcement. This is precisely what is observed in table 2 of Stice's paper: there is a below-average level of trading in each of the five three-day intervals preceding the announcement and an above-average level of trading in each of the four subsequent three-day intervals.[6]

More speculatively, it might be possible to explain a delayed price reaction, such as that observed by Stice, by a model of an information monopolist as in Kyle (1985). If small investors do not find it worthwhile to gather 10-Ks directly from the SEC,[7] and only a single investor is large enough to find it worthwhile to collect the information, that investor will be a temporary information monopolist and will trade in such a manner that the information will be reflected in share prices only gradually.[8]

However, perhaps the most severe challenge to financial theorists posed by recent work on market reactions to earnings announcements is the evidence of Ou and Penman (1989a, b) that the market underutilizes information that is useful in predicting future earnings, and the finding by Bernard and Thomas (1989, 1990) that market reactions to the current quarter's earnings can be predicted on the basis of information contained in the previous quarter's announcement, and that this phenomenon can be explained by a model in which the market uses a naive model of the stochastic process for earnings.[9] Is it possible that modest anomalies can persist unnoticed until brought to light by the high resolution power of the computer operating on a large data base? While it is tempting to believe so, we should be reluctant to accept such explanations until the alternatives have been exhausted.[10] Nevertheless, it is worth noting in this regard that Black and Scholes (1972) found large discrepancies from their model in the pricing of options on the over-the-counter market prior to the development of their model. These discrepancies largely disappeared as their model became the industry standard for pricing in the newly emergent Chicago markets. Similarly, financial markets have reacted to evidence that small firms produce exceptionally large returns by the development of low-cost strategies for investing in such firms.

II. Stock Prices and Valuation

There is a long history of the use of accounting data to explain the relative stock prices of different firms or to estimate the cost of capital for firms in a given industry by cross-section regression of stock prices on measures of earnings and growth. These efforts were largely abandoned by the 1970s because it was found that the regression coefficients were unstable and that deviations from the regression line represented, not mispricing, but rather model misspecification and omitted variables.[11] However, it is still common practice for financial economists to combine accounting data with stock prices to obtain estimates of the cost of equity capital for public utilities for regulatory purposes. The assumption being made is that the dividend payout ratio and the accounting return on equity are constant, and that therefore the growth rate in dividends is equal to the product of the earnings retention ratio and the return on equity.[12] With this exception, the finance profession has moved away from trying to relate stock prices to more fundamental variables, so that by 1987 Ross was able to write that the methodology of finance is the use of close substitutes to price financial assets and contracts. Thus the not inconsiderable empirical successes in finance relate one asset price or return to that of a portfolio of other assets, and, until the recent debate on stock market volatility,[13] empirical researchers had largely ignored the task of relating stock prices to more fundamental variables.[14]

---

[6] The decline in trading prior to the announcement suggests the need to consider variations in volume induced by changes in the level of adverse selection in the market as the result of a public signal, as well the volume effects induced by changing beliefs discussed by Marshall (1974), Hakansson et al. (1982), Varian (1986), and others. The degree of adverse selection in the market associated with public announcements should also be reflected in variation in the bid-ask spread (see Glosten and Milgrom 1985). Barclay and Smith (1988) find such an effect associated with corporate buy-backs.

[7] The infrequency of the delayed *WSJ* announcements observed by Stice suggests that relying on the 10-Ks to beat the *WSJ* announcement is likely to be a costly strategy for the small investor.

[8] However, Holden and Subrahmanyam (1990) show that the speed of price adjustment is highly sensitive to the number of informed investors.

[9] It would be interesting to determine whether analysts' forecasts exhibit the same naive properties.

[10] Evidence of persistent price drift after earnings announcements has been available for more than a decade.

This task has been largely taken over by accounting researchers. Although the majority of studies have focused on the relation between stock prices and accounting earnings, the Forum article by Barth et al. in this issue is concerned with the relation between share prices and certain items that are supplemental to the balance sheet. While it is of obvious interest to determine the extent to which investors take account of accounting measures of loan portfolio quality, it is by no means clear that cross-section regressions of this kind can yield satisfactory answers. First, the instability of regression coefficients across years, referred to above, that is symptomatic of the omission of important variables, is apparent in this study also. A particular reason to believe that omitted variables may be an important consideration in this type of study is the theory of signaling in financial markets,[15] which suggests that managers may choose certain balance sheet configurations (e.g., the amount of debt financing) precisely to communicate to the market their private information about other firm characteristics (e.g., future profitability) that are necessarily omitted from the regression specification.[16] In the present context, it is at least arguable that those thrifts that are in better financial condition might be more willing to make a generous allowance for loan losses than would those thrifts closer to violating regulatory capital requirements. Further, as the authors point out, another omitted variable, the value of the deposit insurance guarantee, will depend on the amount of capital of the thrift and will be related to the right-hand-side variables. There are also some other specifics of the regression model that may appear problematic. First, the regression is specified in terms of per share variables in order to take account of heteroscedasticity. This seems inappropriate because a simple adjustment for heteroscedasticity divides the whole equation (including the constant term) by the deflator, whereas the current study divides everything but the constant. The effect, which may be seen by multiplying the equation by the number of shares, is to introduce the number of shares as a separate independent variable—in this case a highly significant one.[17] There is no reason why the value of a thrift might not depend on the number of shares outstanding—this variable might proxy for size or it might be a signal of management's private information;[18] however, it is clear that introduction of this variable has a potentially dramatic effect on the nature of the regression model and therefore on the interpretation of the regression coefficients.

---

[11] See Keenan (1970) on the "Great SERM Bubble."

[12] See Myers (1972). Note the similarity between this simple model and the "clean surplus" approach of Ohlson (1989b).

[13] This was initiated largely by economists outside the mainstream finance profession such as Shiller (1981).

[14] An unusual exception to Ross's generalization is Chen et al. (1986). Not without some justification, Summers (1985) parodies as "ketchup economists" those within the finance tradition who are concerned only with the interrelationships between the prices of different financial assets.

[15] See J. Williams (1989).

[16] A similar point is made by Beaver et al. (1989).

Second, the specification of the valuation error for non-loan assets as a linear function of the equity book value seems overly simple; very strong assumptions have to be made about the error terms to arrive at a well-specified regression. In summary, it is not clear that this study does not suffer from all of the problems that led to the collapse of the "Great SERM Bubble."

Turning to the extensive literature that attempts to relate stock prices to accounting earnings information, the underlying valuation model, insofar as there is one,[19] is that of Miller and Modigliani (1961). There it was shown that the value of the firm's equity can be expressed either as the present value of the expected future dividend stream to the current shareholders or as the sum of the capitalized current earnings and the present value of future investment opportunities.[20] These two approaches are equivalent because of the dividend irrelevance proposition that follows from the firm's sources and uses of funds constraint. Most accounting researchers, interested in the role of accounting earnings for valuation, have tended to focus on the second of these formulations and have treated accounting earnings as a proxy for the Miller-Modigliani earnings concept. Unfortunately, accounting income does not correspond to the theoretical economic concept of income because the Miller-Modigliani (MM) valuation formula assumes that earnings on the existing assets are defined so that they follow a "uniform perpetual" stream. While the MM formula could be generalized to a situation in which the earnings on the existing assets follow a martingale process, it does not apply (even under certainty) unless the expected change in earnings (excluding the effects of future investments) is zero.

Even supposing that earnings do follow a martingale after adjusting for the effect of new investment, the simple capitalization of current earnings omits the value of future investment opportunities, which, as we have already noted, can bias the coefficients of a regression. Recognizing this problem, researchers[21] have followed the example of Ball and Brown (1968) by focusing on the relation between changes in stock prices and contemporaneous (unexpected) changes in earnings. The implicit assumption in first differencing is that, while levels of the variables missing from the price-earnings relation are correlated with earnings, innovations in these variables are not correlated. This is a heroic assumption in a model with only a single variable included.

---

[17] Note that the number of shares remains statistically significant even when a constant term is introduced into the original (undeflated) regression. See footnote 7.

[18] See Brennan and Hughes (1990).

[19] It has been suggested to me that the purpose of much of this literature is only to determine whether a particular set of variables provides incremental explanatory power for security prices relative to another set of variables, whether this incremental explanatory power arises because market participants are responding to the information or to other value-relevant information to which it is correlated. Such a purpose seems unduly circumscribed, however.

[20] See Miller and Modigliani (1961, 416, eq. 12). This model relies on the *ad hoc* assumption that risk can be taken into account by an appropriate adjustment of the discount rate.

[21] See, for example, Beaver et al. (1980).

The early studies that followed Ball and Brown, using regression techniques, found that the elasticity of stock prices with respect to the unexpected component of annual earnings changes was only about 0.2 instead of unity as a naive model (based on a martingale model of earnings and cash flows) would imply; relatively little attention was paid to the magnitude of the coefficients, so that the analysis that did not go far beyond confirming the association between earnings and stock prices that had been documented by Ball and Brown. Only recently has attention turned to the determinants of earnings response coefficients. I shall argue that the low value of the regression coefficient and the small fraction of variance explained by these cross-section regressions should not be regarded as cause for despair; there are good reasons to expect such results, even if earnings are an important factor in valuation.

I have already observed that the regression of price change on unexpected changes in earnings is likely to be misspecified and to omit important variables that influence expectations about future levels of earnings. Errors in the measurement of the independent variable will also bias down the estimated coefficient. In one of the few studies to acknowledge the consequences of errors in the measurement of unexpected earnings, Brown et al. (1987) demonstrate that the use of instrumental variables in a regression of abnormal returns on unexpected earnings (deflated by the stock price) can raise the coefficient estimate from around unity to about 3.4, at the same time doubling the R^2 of the regression.

In addition to simple errors of measurement, the definition of earnings used in these studies would also seem to be of importance. Yet the careful consideration given by practicing accountants to the manner in which earnings are reported, and to which items are included in earnings, stands in marked contrast to the casual attitude of most researchers toward the definition of the variable under investigation.[22] There are four earnings figures reported on the COMPUSTAT tape, constructed from primary and fully diluted earnings from continuing operations, before and after extraordinary items.[23] However, even earnings before extraordinary items include several items that are of a one-time or exceptional nature, such as losses due to restructuring of operations, asset write-downs, and asset sales. Table 1 reveals that in a given year some 20 to 30 percent of firms report losses due to these one-time events, and even if these losses are not large relative to total earnings, they are likely to be large relative to the change in earnings or unexpected earnings, the variables used in most empirical work.[24] It seems likely that asset write-downs contain considerable information that is useful for forecasting future operating income. For example, insofar as write-downs reflect an acceleration of future depreciation expense one might expect them to be associated with increases in future operating earnings; yet the martingale model of accounting earnings seems to assume that asset write-downs portend future asset write-downs and lower earnings in the future.

---

[22] A similarly cavalier attitude toward the definition of the market portfolio is to be found in early studies testing the capital asset pricing model.

[23] Presumably the best approach to dealing with potential dilution is to value all the equity first and to deduct the value of convertible securities using a contingent claims approach to valuation.

[24] For example, Bethlehem Steel reported an increase in income from operations of 468 percent in 1988; this was reduced to 260 percent if account is taken of the one-time nature of gains and losses on discontinued businesses and asset sales. The operating income growth of American Brands in 1987 was a healthy 34 percent; but taking account of the fact that 1986 earnings were affected by a major restructuring charge reduces this growth to only 13 percent.

Table 1

Number of Companies out of 600 Disclosing Various Losses that Affect Income from Continuing Operations

Restructuring of Operations	74	81	133	103
Foreign Currency Transactions	96	78	96	114
Write-down of Assets	43	30	58	49
Sale of Assets	25	28	27	29
Litigation Settlements	17	11	11	17
Total[1]	255	228	325	312

Source: *Accounting Trends and Techniques* (1990)

[1] Includes companies that report more than one type of loss.

In regressing abnormal stock returns on (unexpected) earnings changes, it is necessary to deflate the right-hand-side variable to make the equation dimensionally consistent. The two popular choices are the current (or expected) level of earnings, and the current stock price. The former deflator essentially assumes that stock prices are proportional to the current level of earnings, whereas the latter embodies the weaker assumption that the stock price is linear in current earnings. The proportionality assumption is certainly problematic since it implies that stock prices will be negative if earnings are, which would be appropriate if there were no limited liability, earnings followed a martingale, and dividends were proportional to earnings. The weaker linearity assumption is also appropriate if there is no limited liability and dividends are proportional to earnings that follow an AR(1) process. While the linearity assumption is to be preferred to that of proportionality, it also is unsatisfactory in that it allows the possibility of negative stock prices. Thus, a significant weakness of much of the literature attempting to relate accounting earnings to stock prices is the lack of an adequate theoretical framework for relating stock prices to accounting earnings.

As Ohlson (1989a) has rightly emphasized, valuation models must at a minimum be consistent with the basic no-arbitrage property of security markets, which implies that stock prices be related to expected discounted dividends. In an important article, Garman and Ohlson (1980) showed that, in a stationary Markovian environment in which the state dynamics can be described by a general linear process, the value of the firm (expected future dividends) can be written as a linear function of the state variables. While Garman and Ohlson do not identify the state variables with particular accounting variables, their analysis suggests that such an identification will be appropriate if accounting variables are to be important for valuation.

The importance of the stochastic process of dividends and the role of earnings in this process unfortunately has been largely ignored in most subsequent studies, which, while paying lip service to a discounted cash flow model of valuation, are extremely casual about the link between such a model and their empirical formulation. Almost always the empirical formulation simply assumes that the stock price is either proportional to, or linearly related to, earnings, and rarely do researchers take advantage of the fact that their independent variable represents earnings, rather than some other income statement variable such as sales, to frame hypotheses about the magnitude of the regression coefficient.

To illustrate, suppose, as Ball and Watts (1972) did, that accounting earnings follow a martingale and that the capitalization rate is ten percent. Then the theoretical earnings response coefficient will depend upon the dividend payout ratio. If the dividend payout ratio is 50 percent, then an unexpected increase in earnings of \$1 will translate into an unexpected increase in dividends of \$0.50 in perpetuity, and this in turn would imply a stock price increase of \$5, or five times the earnings increase. Suppose, on the other hand, that earnings are *iid*. If the dividend payout rate is 50 percent, then the abnormal return associated with the \$1 unexpected increase in earnings is only \$0.50. More complex calculations are required when the earnings process is more general. However, the point is that there is little to be gained from estimating an earnings response coefficient unless it is to test hypotheses framed in terms of the *joint* stochastic process of earnings and dividends.

Viewed in this light, most estimates of earnings response coefficients are distressingly low, given the conventional wisdom that earnings are well-approximated by a martingale. Cornell and Landsman (1989), for example, found a value of around 1.3 for the coefficient on the revision in analysts' forecasts of the following year's earnings. This is more consistent with a model of *iid* earnings than with a martingale, unless the dividend payout is very low.

Beaver et al. (1980) recognize the importance of the assumption about the stochastic process for earnings, but they do not allow for interfirm differences and are not explicit about the dependence of dividends on earnings. In an important paper, Kormendi and Lipe (1987) use a rational expectations framework to model the dependence of the earnings response coefficient on the stochastic process for earnings. They find an association between the response coefficient and the stochastic process as predicted. Unfortunately, the valuation model they employ implicitly assumes that the whole of each firm's earnings are paid out in dividends, and this may be the reason for their formal rejection of the model.[25] Nevertheless, this paper is significant in presenting a formal test of the relation between the stochastic process for earnings and the stock price response to earnings changes.

III. Alternative Approaches

I have criticized earlier studies for not articulating the relation between earnings and dividends and for not distinguishing carefully enough between economic and accounting earnings concepts; indeed, the existing empirical studies pay no heed to the essential feature of accounting earnings, that earnings not paid out in dividends increase the book value of equity. This accounting convention corresponds to the economic reality that an unexpected increase in earnings, if not matched by an increase in dividends,[26] will increase the net assets available to generate future earnings and pay dividends. Therefore, in order to have a valuation model in which accounting earnings play a role, it is necessary to consider the balance sheet as well as the income statement. Brennan and Schwartz (1982, 1984b) develop a simple model along these lines. The basic assumptions are that:

(i) x, the return on equity, follows a simple mean reverting process

$$dx = a(m - x)dt + s\ dz,$$

---

[25] See Kormendi and Lipe (1987, fn. 33).

[26] And assuming that it corresponds to an unanticipated cash flow.

where dz is the increment to a Gauss-Wiener process; m is the long run mean return on equity; a is the speed of adjustment; and t is calendar time. The motivation for this assumption is that unusually high or low earnings in relation to equity values are unlikely to persist, as the company or its competitors will take actions that will tend to restore profitability to a normal level.[27]

(ii) The total dividend payout is a function of the book value of the equity, B, and the return on equity $D(x,B)$, that can be written in the form

$$D(x,B)=p(x)B,$$

where $p(x)$ is the payout ratio. This is an obvious simplification that is intended to make the model tractable. More elaborate specifications are possible.

(iii) The rate of change in book equity is equal to the difference between the current earnings rate xB and the payout rate $p(x)B$:

$$dB=(x-p(x))B\ dt.$$

This is the "clean surplus" condition of Ohlson (1989b); it does not allow for asset write-downs or increments to equity by stock sales.

It can then be shown that the market value of the firm's equity can be written as the product of the book value of the equity and a market to book ratio that depends only on the return on equity:

$$E(x,B)=v(x)B.$$

The market to book ratio, $v(x)$, is a nonnegative increasing function of the return on equity that depends on the precise terms on which the equityholders can declare bankruptcy. In this very simple model, accounting earnings do not follow a martingale, so that accounting earnings is not synonymous with economic earnings and the stock price is not a linear function of the earnings; yet accounting earnings, together with the current book value, are sufficient statistics for the value of the shares. An important aspect of the model is that the relation between the stock price and the earnings depends both on the current rate of return on equity and on the assumed stochastic process for the return on equity, as well as a risk adjustment term and the dividend payout policy. As in Garman and Ohlson (1980), there are consistency relations between the assumed dynamics of the state variables, x and B, and the function relating the stock price to the state variables, in this case the $v(x)$ function. This is a very simple model because of the simple characterization of the return on equity process; however, it has the advantage of explicitly accounting for the limited liability character of corporate equity.

A second approach to modeling the relation between stock prices and accounting variables, which takes proper account of the role of dividends but does not consider the possibility of bankruptcy, is the recent work of Ohlson (1989b) that develops the linear information dynamics model of Garman and Ohlson (1980). Thus, define E_t, B_t, D_t, and P_t as the earnings, book value, dividends, and price per share at time t. Then the Ohlson approach may be represented by the following equation system:

$$E_{t+1}=a_{11}E_t+a_{12}B_t+a_{13}D_t+v_{1t} \tag{A-1}$$

---

[27] Sometimes this will be accomplished by asset write-downs. These are easily included in our model although we do not include them here.

$$B_{t+1} = a_{21}E_t + a_{22}B_t + a_{23}D_t + v_{2t} \tag{A-2}$$

$$D_{t+1} = a_{31}E_t + a_{32}B_t + a_{33}D_t + v_{3t} \tag{A-3}$$

$$P_{t+1} = \quad cP_t + \quad D_t \qquad + v_{4t}, \tag{E}$$

where v_{it} are random errors. The first three equations describe the evolution of earnings, book value, and dividends, and the coefficients of the equation (A-2) are constrained by the "clean surplus" relation that the change in book value be equal to the retained earnings. Equation (E) is the equilibrium condition that the expected rate of return on the stock be equal to $(c-1)$; in Ohlson's model investors are assumed to be risk neutral so that c is the gross riskfree interest rate that is assumed to be constant.[28] The linear structure of the system (A) means that the expected dividend at time $(t+j)$ can be expressed as a linear function of the current values, B_t, E_t, D_t, and the stock price, P_t, which is equal to the discounted sum of these dividend expectations, can also be expressed as a linear sum of the same variables:

$$P_t = b_1E_t + b_2B_t + b_3D_t + v_{4t}, \tag{B}$$

where the coefficients b_j are functions of the coefficients of the systems A and E. Thus, estimation of the equation system A, E, and B taking account of the coefficient restrictions yields a test of the hypothesized relation between accounting variables and stock prices.

A somewhat related approach is implicit in the work of Campbell and Shiller (1988), who show that rational valuation implies restrictions on the coefficients of a vector autoregression involving the log dividend yield, the interest rate, and the growth rate of dividends. Such an approach should lend itself to investigation of the relation between stock prices, dividends, and accounting variables at the individual firm or industry level. Share prices are related to expected future dividends; if accounting numbers carry information about future dividends, then rational valuation implies restrictions on the relation between stock prices and accounting numbers.

In summary, while disappointment has been expressed at the lack of progress made since 1968 in understanding the relation between accounting and valuation, considerable progress has been made, and the prospects for further progress are good if future research pays careful attention to the economics of the relation between the fundamentals represented by accounting information and share prices.

---

[28] In empirical implementation c could be allowed to be a function of the current riskless rate and the state variables, E_t, B_t, and D_t. The equation system (A) would then have to be augmented by a fourth equation describing the evolution of the riskless interest rate.

References

Admati, A. R., and P. Pfleiderer. 1988. A theory of intraday patterns: Volume and price variability. *Review of Financial Studies* 1 (Spring): 3–40.

Atiase, R. K. 1985. Predisclosure information, firm capitalization and security price behavior around earnings announcements. *Journal of Accounting Research* 23 (Spring): 21–36.

Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6 (Autumn): 159–78.

———, and R. Watts. 1972. Some time series properties of accounting income. *Journal of Finance* 27 (December): 663–82.

Barclay, M. J., and C. W. Smith. 1988. Corporate payout policy: Cash dividends versus open market repurchases. *Journal of Financial Economics* 22 (October): 61–82.

Barth, M. E., W. H. Beaver, and C. H. Stinson. Supplemental data and the structure of thrift share prices. *The Accounting Review*, this issue.

Beaver, W. H. 1968. The information content of annual earnings announcements. *Journal of Accounting Research* 6 (Supplement): 67–92.

———, C. Eger, S. Ryan, and M. Wolfson. 1989. Financial reporting, supplemental disclosures and the structure of bank prices. *Journal of Accounting Research* 27 (Autumn): 157–78.

———, R. A. Lambert, and D. Morse. 1980. The information content of security prices. *Journal of Accounting & Economics* 2 (March): 3–28.

Bernard, V. L., and J. K. Thomas. 1989. Post earnings announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27 (Spring): 1–35.

———, and ———. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. Unpublished manuscript, University of Michigan.

Black, F., and M. S. Scholes. 1972. The valuation of option contracts and a test of market efficiency. *Journal of Finance* 27 (May): 399–417.

Brennan, M. J., and P. J. Hughes. 1990. Stock prices and the supply of information. Unpublished manuscript, University of Southern California.

———, and E. S. Schwartz. 1982a. Consistent regulatory policy under uncertainty. *The Bell Journal of Economics* 13 (Autumn): 506–21.

———, and ———. 1982b. Regulation and corporate investment policy. *Journal of Finance* 37 (May): 289–300.

Brown, L. D., P. A. Griffin, R. L. Hagerman, and M. E. Zmijewski. 1987. An evaluation of alternative proxies for the market's assessment of unexpected earnings. *Journal of Accounting & Economics* 9 (July): 159–93.

Campbell, J. Y., and R. J. Shiller. 1988. The dividend price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies* 1 (Fall): 195–228.

Chen, N-F., R. Roll, and S. A. Ross. 1986. Economic forces and the stock market. *Journal of Business* 59 (July): 383–404.

Cornell, B., and W. R. Landsman. 1989. Security price response to quarterly earnings announcements and analysts' forecast revisions. *The Accounting Review* 64 (October): 680–92.

Diamond, D. W., and R. E. Verrecchia. 1987. Constraints on short selling and asset price adjustment to private information. *Journal of Financial Economics* 18 (June): 277–311.

Foster, D. F., and S. Viswanathan. 1990. Variation in volumes, spreads and variances. *Review of Financial Studies*. Forthcoming.

French, K. R., and R. Roll. 1986. Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics* 17 (September): 5–26.

Garman, M., and J. A. Ohlson. 1980. Information and the sequential valuation of assets in arbitrage-free economies. *Journal of Accounting Research* 18 (Autumn): 420–40.

Glosten, L. R., and P. R. Milgrom. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14 (Autumn): 71–100.

Grundy, B. D., and M. McNicholls. 1989. Trade and revelation of information through prices and direct disclosure. *Review of Financial Studies* 2 (Winter): 495–526.

Hakannsson, N. H., J. G. Kunkel, and J. A. Ohlson. 1982. Sufficient and necessary conditions for information to have social value in pure exchange. *Journal of Finance* 37 (December): 1169–81.

Hellwig, M. F. 1980. On the aggregation of information in securities markets. *Journal of Economic Theory* 22 (June): 477–98.

Holden, C. W., and A. Subrahmanyam. 1990. Long-lived private information and imperfect competition. Typescript, University of California, Los Angeles.

Keenan, M. 1970. Models of equity valuation: The great SERM bubble. *Journal of Finance* 25 (May): 243–73.

Kormendi, R., and R. Lipe. 1987. Earnings innovations, earnings persistence, and stock returns. *Journal of Business* 60 (July): 323–46.

Kyle, A. S. 1985. Continuous auctions and insider trading. *Econometrica* 53 (November): 1315–35.

Lev, B. 1988. How useful are earnings? Two decades of empirical findings. Typescript, University of California, Berkeley.

Marshall, J. 1974. Private incentives and information. *American Economic Review* 64 (June): 373–90.

Miller, M. H., and F. Modigliani. 1961. Dividend policy, growth and the valuation of shares. *Journal of Business* 34 (October): 411–33.

Myers, S. C. 1972. The application of finance theory to public utility rate cases. *The Bell Journal of Economics and Management Science* 3 (Spring): 58–97.

Ohlson, J. A. 1989a. Ungarbled earnings and dividends: An analysis and extension of the Beaver, Lambert and Morse valuation model. *Journal of Accounting & Economics* 11 (April): 109–15.

———. 1989b. Accounting earnings, book value, and dividends: The theory of the clean surplus equation. Typescript, Columbia University.

Ou, J. A., and S. H. Penman. 1989a. Financial statement analysis and the prediction of stock returns. *Journal of Accounting & Economics* 11 (Autumn): 295–329.

———, and ———. 1989b. Accounting measurement, price-earnings ratios, and the information content of security prices. *Journal of Accounting Research* 27 (Spring): 111–44.

Roll, R. 1988. R^2. *Journal of Finance* 43 (July): 541–66.

Ross, S. A. 1987. Finance. In *The New Palgrave Dictionary of Economics*, edited by J. Eatwell, M. Milgate, and P. Newman. London: Macmillan.

Shiller, R. J. 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71 (June): 421–36.

Stice, E. 1991. The market reaction to 10-K and 10-Q filings and to subsequent *The Wall Street Journal* earnings announcements. *The Accounting Review*, this issue.

Summers, L. H. 1985. Economics and finance. *Journal of Finance* 40 (July): 633–36.

Swaminathan, S. 1991. The impact of SEC mandated segment data on price variability and divergence of beliefs. *The Accounting Review*, this issue.

Varian, H. R. 1986. Differences of opinion in financial markets. Typescript, University of Michigan.

Wiang, J. 1989. Asset prices, stock returns, price volatility, risk premium and trading strategies under asymmetric information. Typescript, Wharton.

Williams, J. 1989. Signalling with dividends, investment and debt. In *Financial Markets and Incomplete Information*, edited by S. Bhattacharya and G. M. Constantinides. Rowman and Littlefield.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 80–99

Budgetary Participation, Locus of Control, and Mexican Managerial Performance and Job Satisfaction

Veronique Frucot
Syracuse University
Winston T. Shearon
Texas A&M University

SYNOPSIS: Researchers generally agree that culture affects the behavior and attitudes of individuals within organizations (Hofstede 1980; Adler et al. 1986). Therefore, it may be assumed that an organizational system such as a participative budget-process will have different effects in different cultures.

Recent research has indicated that a personality variable, internal-external locus of control, impacts the relationship between budgetary participation and both managers' performance and job satisfaction (Brownell 1981, 1982b). The locus of control construct categorizes individuals as (1) externals, those who believe that events are controlled by fate, luck, chance or powerful others, or (2) internals, those who believe that they have some control over events (Rotter 1966). This article examines whether cultural differences affect the previously identified interrelationship of individual locus of control and participation in the budgeting process as it impacts managerial performance and satisfaction.

The prior research (Brownell 1982b) used middle-level managers in U.S. manufacturing organizations. In this study, Mexican managers were selected because Mexico provides an interesting cultural contrast to the U.S. on three dimensions identified by Hofstede (1980) which are considered relevant to participative budgeting issues. In contrast to the "Anglo" Cluster, the Latin Cluster of which Mexico is a unit differs on the dimensions of "uncertainty avoidance" and "power distance." An addi-

The authors acknowledge the helpful comments of Mohamed Onsi, Robert Strawser, William Wasserman, the editor, two anonymous referees, and especially John C. Anderson. In addition, special thanks are due to Lic. Guillermo Torrescano and Monique Frucot for their assistance in contacting firms.

Submitted May 1989.
Revisions received March and August 1990.
Accepted August 1990.

tional factor in the selection of Mexican managers was the magnitude of economic ties between Mexico and the U.S.[1] The importance of this relationship continues to grow as more multinationals establish "maquilladoras"[2] within the border region and as the Mexican government frees some of its constraints on foreign investment.

The responses of 83 Mexican managers to survey instruments were analyzed using regression to test the interrelationship of locus of control and budgetary participation and their impact on managerial performance and job satisfaction. While the results of this study are generally consistent with previously reported findings for managerial performance, the impact of locus of control on managerial satisfaction was not significant, reflecting an ostensible difference in culture. In addition, the effect of locus of control on the performance of high-level managers was significantly stronger than its impact on the performance of lower-level managers. Finally, contrasting significantly to other Mexican managers, the performance of those Mexican managers employed in 100 percent foreign-owned firms was not discernibly affected by either budgetary participation or locus of control. This last result may be due to the cognitive dissimilarities relating to the cultural interface in these foreign-owned firms. This suggests that, within the Mexican culture, different conclusions are obtained depending on whether the firm is controlled by local or foreign interests.

Key Words: *Participative budgeting, Managerial performance, Managerial satisfaction, Locus of control, Mexico.*

Data Availability: *A hard copy of the summarized data used in this study may be obtained from the first author on request.*

THE effectiveness of the management planning and control systems used in developing countries remains an important issue for a number of reasons. First, as developing countries become more industrialized, progressive local firms may import systems and techniques from the more industrialized nations; indeed, such importation may be encouraged by educational institutions and consultants. Unfortunately, the progressive local firm may be importing systems from the industrialized world without considering the potential impact of cultural differences on the success of implementing such systems.

---

[1] Mexico is the second largest borrower from U.S. banks (Riding 1985, 149; Gilbreath 1986, 7), the largest foreign supplier of crude oil to the United States (Riding 1985, 334), the third largest trading partner of the United States (Riding 1985, 333; Gilbreath 1986, 4), and the fourth largest importer of goods from the United States (Gilbreath 1986, 5). In addition, many of the *Fortune 500* companies have invested directly in Mexico. This trend may be increasing since the Mexican government has relaxed certain of its limitations on foreign capital. As reported by Gilbreath (1986, 13–14), these factors and a common border have led to a strong economic relationship between the two countries.

[2] Maquilladoras are across-the-border assembly plants that use imported parts and supplies and reexport their output to the U.S. or other countries. These plants in the border area are currently exempt from the requirements of Mexico's foreign investment laws. While this manufacturing arrangement was actually devised about 25 years ago, the expansion of maquilladoras began in earnest in 1982 following the devaluation of the Peso. Many large U.S., European, and Japanese firms have been attracted by Mexico's low labor costs. The number of maquilladoras in Northern Mexico reached 1,250 in 1986, employing more than 300,000 workers (Baker et al. 1988).

Second, multinational firms are faced with an additional problem arising from the multicultural setting in which they operate; the interaction of organizational members with diverse cultural backgrounds. This cultural interface can occur at various levels within the organization; for example, between the subsidiaries' high- and low-level management (i.e., when an American executive is sent to manage the Mexican subsidiary) or between the management of the subsidiary and that of the parent. The diverse cultural backgrounds translate into a lack of cognitive similarities which can significantly hamper communication at that interface (Stening 1979).

Support for the concern about the issues outlined above is found in current definitions of culture as a "shared reality" or shared "cognitive maps" (Diaz-Guerrero 1975, 1984; Adler et al. 1986; Triandis 1972; Hofstede 1980, 1987). Through the socialization process, members of a group learn the premises that make up this "subjective culture." These premises provide the essential basis for interpersonal relations leading to the effects described by Adler et al. (1986, 299):

> Culture influences organizations through societal structures such as laws and political systems and also through the values, attitudes, behavior, goals, and preferences of participants.

Thus, the view of "reality" held by members of a group is expected to affect their perception of the budgetary process and the values that they place on participation in that process (Hofstede 1987).

I. Budgeting, Participation, and Locus of Control

The behavioral consequences of participative budgeting, particularly performance, job satisfaction, motivation, and dysfunctional behavior, have been extensively investigated. Research results indicate that participation has a positive effect on managerial satisfaction in the United States (Milani 1975; Collins 1978; Hofstede 1967; Cherrington and Cherrington 1973; Swieringa and Moncur 1975; Brownell 1982b; Chenhall 1986). However, the relationship between participative budgeting and managerial performance is less clear. While some research has indicated a strong positive relationship (Argyris 1952; Brownell 1982b; Hofstede 1967; Kenis 1979; Merchant 1981; Brownell and McInnes 1986), other findings suggest a weak or negative relationship (Milani 1975; Cherrington and Cherrington 1973; Bryan and Locke 1967; Locke and Schweiger 1979).

Various researchers have suggested that the relationship between participative budgeting and performance might be contingent upon the presence of other moderating factors. The moderating factors that have been tested can be grouped into (1) macrolevel variables, such as organizational structure and technology, and (2) microlevel variables, such as leadership style and individual personalities (for a review, see Brownell 1982a). Across cultures, convergence is most likely to be found at the macrolevel since such variables are not affected as much by cultural differences as they are by industry, size, and market factors (Child 1981; Child and Tayeb 1983; Levitt 1983; Garnier 1984; Adler 1983; Adler et al. 1986).

Cultural diversity may, however, play an influential role at the microlevel (Adler 1983; Adler et al. 1986; Hofstede 1980, 1983; Inzerelli and Laurent 1982–1983; Laurent 1983; Daley et al. 1985; Kanungo and Wright 1981). With respect to participative budgeting, managerial locus of control, which is a microlevel moderating variable, has been investigated with consistent findings in the U.S. (Brownell 1981, 1982b; Licata et al. 1986). The results of these studies indicate that "internals" perform better and are more satisfied than "externals" when the budgetary process is participatory (Brownell 1981, 1982b) and that internal supervisors or managers are more participative in their budget-setting style (Licata et al. 1986). These results are consistent with the findings of Runyon (1973), Mitchell et al. (1975), and Goodstadt and Hjelle (1973). An additional consideration in the selection of the locus of control construct was its relevance to cultural settings other than the United States, including Mexico (for reviews, see Lefcourt 1976, 1982 and Phares 1976).

Hofstede (1980, 1983) investigated microlevel cultural differences in the work-related value patterns of individuals employed by a large multinational company in 50 countries and three regions in 1968 and 1972. Hofstede used respondents who were well-matched in both their occupational categories and education level. Since all were employed by the same company, they shared the same overall company structure and policies. Although they differed as to age and gender, the only systematic difference was nationality. He found that half of the variation in a country's mean score could be explained by four dimensions:

- *Power distance:* a measure of the degree to which cultures prefer a more autocratic structure. In an organizational context, it translates into the degree of participation that would be favored. A country scoring high on this construct would prefer a less participative style.
- *Uncertainty avoidance:* a measure of the mean anxiety level. A high anxiety level translates into less willingness to take risks and a preference for security. Organizations in a high uncertainty avoidance environment would favor rigid rules, so as to decrease uncertainty.
- *Individualism:* a measure of the relative importance of independence from the organization. Cultures scoring high on this construct stress goals and independence. Cultures scoring low on this construct favor more dependence on the organization.
- *Masculinity:*[3] a measure of the relative importance to the culture of income, recognition, and advancement as compared to the importance of work relations, cooperation, and security.

As indicated by Hofstede, these dimensions constitute a framework for developing hypotheses in cross-cultural studies of organizations. Indeed, he calculated an index score on each of the dimensions for each country, and validated his results by comparing them to other samples and to country level indicators.

Based on these dimensions, respondents from the United States and Mexico would be classified as follows:

	U.S.	*Mexico*
Power distance	Low	High
Uncertainty avoidance	Low	High
Individualism	High	Low
Masculinity	High	High

---

[3] In his research, Hofstede used the term "masculinity" for this dimension. While the dimension itself is relevant to this study, the exact descriptive term is not.

The potential implications of this classification are:

1. In the United States, a more participative decision-making style would be preferred resulting in less centralization of decision making at the top level of the organization and fewer rules. Self-actualization and achievement are emphasized over social needs, such as belonging. In addition, an individual would prefer to be less dependent on the organization.
2. In Mexico, a more paternalistic and autocratic leadership style (with less participation in decision making and more rules and centralization) would be preferred. In addition, more dependence of the individual on the organization is expected. However, consistent with the United States, achievement and self-actualization are to be emphasized over more collective values.

Based on Hofstede's classification scheme, we would expect Mexican managers to generally favor a less participative budgeting system than would their American counterparts. Thus, the relationship between participation in the budgeting process and managerial satisfaction may be different in Mexico than in the U.S. However, it is important to note that Hofstede directed his questionnaires to employees rather than focusing specifically on managers who, according to Brownell (1982b), tend to be more internal.

It is difficult to predict the effect of participation on the budgeting process for Mexican managers in general. While research in budgetary participation has demonstrated the importance of locus of control as a moderator between participation and performance-satisfaction, cross-cultural management studies have emphasized cultural dimensions as determinants of the type of organization preferred. Although the Mexican managers who are classified as internals would be expected to prefer a more participative budgeting style, cultural influences could lead these same managers to be more amenable to both an autocratic structure and a certain degree of dependence on the organization (positions which are more in line with an external personality) than would their U.S. counterparts.

If Hofstede's premise regarding the preference for a more autocratic structure in Mexico is correct, we would expect Brownell's model to apply mostly to higher-level managers. Because of their high position in the hierarchy of their organizations, these managers are expected to assume a leadership role and would be more involved in the participatory budgeting process. As we move down the managerial levels, however, the model's applicability would be expected to decline due to the power distance and uncertainty avoidance characteristics described by Hofstede. Thus, we would expect a difference in the effects of budgetary participation on high-level versus low-level managers.

Another factor is the multicultural setting in which the subsidiaries of multinationals operate. Cross-cultural research supports the view that cross-cultural communication is hampered by a lack of cognitive similarities (for a review, see Stening 1979). Individuals who view their environment in a similar fashion will perceive their communication as more effective than when the parties involved are cognitively different (Triandis 1960). Similarly, supervisors and subordinates like each other more when they are cognitively similar (Triandis 1959). Thus, within the same culture, cognitive dissimilarities between individuals reduce communication effectiveness. According to Stening (1979), this problem is exacerbated in intercultural communication where cognitive dissimilarity is greater. This conclusion implies that multinationals with foreign subsidiaries would experience greater communication problems within their organization than comparable single-culture firms. Since the budgetary process is part of a firm's internal communication system, cognitive dissimilarities could influence its effectiveness. However, the specific expected effect of this problem is unclear.

II. Hypotheses and Test Instruments

Adler (1984) stresses that it is important to replicate research performed in one national culture, in order to increase our knowledge of culture's impact. To facilitate comparability, Adler also recommends that the replication in another culture follow the initial study's research method as closely as possible. Therefore, the first four hypotheses tested in this study follow Brownell (1982b).

Traditionally, it is hypothesized that participation will lead to both higher performance and increased satisfaction. In addition, from the previous budgeting literature, it is anticipated that an internal locus of control will accentuate that relationship, while an external locus of control will attenuate it. Thus, the first four hypotheses in their null form are as follows:

Ho1: In Mexico, the degree of participation in the budgetary process does not lead to higher performance.
Ho2: In Mexico, the degree of participation in the budgetary process does not lead to higher satisfaction.
Ho3: In Mexico, locus of control does not have, through an interaction effect with budgetary participation, a significant effect on performance.
Ho4: In Mexico, locus of control does not have, through an interaction effect with budgetary participation, a significant effect on satisfaction.

If this study examined similar managers in the same culture as did previous studies, for example, Brownell (1982b), we would expect similar results. However, the conflict discussed earlier between the effects of internal locus of control and cultural differences could lead to different findings. In particular, the high power distance and high uncertainty avoidance, which Hofstede (1980, 1987) found to be prevalent in Mexico, should lead to preference for less participation in the budgeting process. Therefore, different results for managerial satisfaction are expected. Managerial performance could also be affected.

Additional specific hypotheses in their null form germane to this study are:

Ho5: In Mexico, management level does not affect the models tested previously (Ho1–Ho4).
Ho6: In Mexico, percentage of foreign ownership does not affect the models tested previously (Ho1–Ho4).

The previous discussion points to a concern for a management-level effect on the results. A prevailing preference in Mexico for an autocratic organization and for governing rules is evident in the high level of power distance and uncertainty avoidance documented by Hofstede. Because of their position and status, higher-level Mexican managers may react to budgetary participation substantially differently than their lower-level counterparts.

The concern for percentage of foreign ownership in the employing firm follows from the potential cognitive difficulties created by the multicultural interface within these firms. As indicated earlier, evidence exists to indicate that American and Mexican managers view organizations differently; thus, communication difficulties between the two groups may affect the participatory budgeting process. Given that these cognitive difficulties are important, we would anticipate that managers in firms totally owned by foreign interests will experience greater difficulties in applying participatory budgeting techniques than will their counterparts employed by firms owned and/or controlled by Mexicans.

Given the concern for the generalizability of inferences derived from field study findings, this study also considers the potential effects of differences in industry and functional area.

The null hypotheses are:

Ho7–Ho8: In Mexico, there is no industry (Ho7), or department effect (Ho8), which would limit the findings pertaining to each of the first four hypotheses.

To test the first four hypotheses, measurements of performance, job satisfaction, budgetary participation, and locus of control were obtained. The specific questionnaires or instruments selected for this study are:

1. Mahoney et al. (1963) self-rating performance questionnaire, an eight-item scale.
2. Weiss et al. (1967) Minnesota Satisfaction Questionnaire, short form, a 20-item eight-point Likert type scale.
3. Milani's measure of participation instrument (1975) which is a six-item, seven-point Likert type scale.
4. Mexican locus of control scale, La Rosa (1986), a 61-item, five-point Likert type scale.

The first three instruments were also used by Brownell (1982b), which facilitates comparisons, and have been tested extensively. These instruments were given in Spanish to each participant. Since the first three instruments were previously available only in English, they were translated by one of the authors and pilot-tested using a method suggested by Hui and Triandis (1985).[4]

The performance questionnaire is based on the results of a survey conducted by Mahoney et al. (1963) that measures eight performance dimensions (planning, investigating, coordinating, evaluating, supervising, staffing, negotiating, representing) and also provides an overall measure of performance. A questionnaire was later constructed based on these dimensions by Heneman (1974) and has been used by Brownell (1982b, 1983), Brownell and Hirst (1986), Brownell and McInnes (1986), Dunk (1989), and others. The question of whether superior or self-ratings are more appropriate has been debated at length.[5] Performance self-ratings were used in this study in lieu of superior ratings.

---

[4] This method consists of having subjects fluent in both the original language of the questionnaire and the language of the translation answer both versions. The logic for this procedure is that if the two versions are equivalent, the subjects should answer each question in exactly the same way. If different answers are given to the same question in different languages, it indicates that the question carries a different meaning in each language and is, therefore, not equivalent and requires that the translation be corrected. Five bilingual subjects, all graduate students, responded to both English and Spanish versions of the questionnaires. Their responses were then compared and checked for inconsistencies and mistakes. The errors in the questionnaires, predominantly spelling errors, were then corrected.

There are two versions of the Minnesota Satisfaction Questionnaire (MSQ): a long form (100 questions) and a short form (20 questions). The short form was used in this study to keep the overall questioning of the subjects within a reasonable time frame. Used in the accounting literature by Chenhall (1986) and Chenhall and Brownell (1988), the short-form version provides three scales of job satisfaction: intrinsic, extrinsic, and an overall measure of satisfaction.

Milani's participation questionnaire measures both the amount of influence an individual has on the budget and his or her involvement in the process of establishing the budget. This questionnaire has been used extensively to study participation in budgeting research (Milani 1975; Brownell 1982b, 1983; Brownell and Hirst 1986; Brownell and McInnes 1986; Chenhall 1986; Chenhall and Brownell 1988; Dunk 1989; Mia 1988, 1989).

La Rosa's questionnaire is divided into five different scales, with each scale measuring a distinct aspect or dimension of internality/externality. The first dimension, internality, measures the extent to which respondents believe that they control their own lives through their own skills and efforts.[6] Three additional scales measure the extent to which other factors control the subject's life:

- Fatalism/luck.
- Powerful figures and institutions distant from the individual, but whose actions still have repercussions on the individual's life (e.g., politicians or government officials).
- Powerful persons who are part of the respondent's immediate surroundings (bosses, supervisors, and relatives).

Finally, the fifth dimension measures the extent to which the respondent believes that his or her goals can be obtained indirectly by being liked by others.

We selected La Rosa's scale because the Rotter (1966) instrument used by Brownell (1981, 1982b) has been criticized for its inapplicability to other groups or cultures such as U.S. Blacks and Hispanics (Gurin et al. 1969; Hui 1982; Hui and Triandis 1983), Mexicans (Hui and Triandis 1983), and Japanese (Weisz et al. 1984). With subjects of a Mexican culture, the sociopolitical segment is problematic primarily because Rotter's scale is unidimensional. In particular, there is a tendency to misclassify internal Mexi-

---

[5] Superior-ratings were used by Mia (1988, 1989) in a budgeting context and some researchers have argued that self-ratings tended to be more lenient (higher mean values) and showed lower variability than did superior-ratings (Parker et al. 1959; Prien and Liske 1962; Thornton 1968). Other researchers have found that this disadvantage did not apply under research conditions (Miner 1968; Heneman 1974). In addition, halo errors (intercorrelations among performance subdimensions) have been found to be more common for superior-ratings than for self-ratings (Prien and Liske 1962; Williams and Seiler 1973; Heneman 1974). Heneman's (1974) results are of particular importance to the present study because it is based on the same instrument. His results indicate that, at least for this instrument, self-ratings are less lenient, and are less subject to halo effect.

[6] The La Rosa questionnaire covers all facets of a respondent's life (job, economy, family, and society in general). As such, some (three) of the questions relate to a student situation. This also is the case for other generalized locus of control questionnaires including Rotter's. Indeed, only one of the student situation questions affects the internality scale and does not unduly impact the measure. Further, any modification of the questions would present potential reliability problems caused by creating a new, untested version of the instrument.

Table 1
Descriptive Statistics Pertaining to the Mexican Managers

Variable	N*	*Mean*	*Possible Range*	*Actual Range*	*Median*	*Standard Deviation*
Age	81	37	—	23–54	37	6.88
Time in position	82	4	—	0.33–15	2.25	3.50
Time in company	81	6	—	0.25–20	5	4.93
Percentage salary increase	66	54	—	0.0–100	56	23.61
Participation**	79	18.21	6–42	6–40	17	7.34
Overall satisfaction**	82	77.43	20–100	21–96	80	12.79
Overall performance**	80	6.99	1–9	4–9	7.05	1.18
Internality**	83	4.36	1–5	3–5	4.33	0.45

* *N*=number of valid responses.
** Participation was measured using a translated version of Milani's participation questionnaire; overall satisfaction was measured using a translated version of the Minnesota Satisfaction Questionnaire; overall performance was measured using a translated version of Mahoney's performance questionnaire; and internality was measured using La Rosa's Internality scale.

can subjects as externals because their sociopolitical environment is reflected in their responses. An overall criticism of Rotter's scale is that it is biased toward white American middle-class values and views.

In addition to these instruments, demographic data were obtained: age, position in the company, length of time position held, level of education, and location of educational facility. An additional question asked for the percentage salary increase received by the manager in the past year.[7] This last item was initially intended as a check on the results obtained with the self-reported performance measure. Since it was not possible to obtain an actual externally-determined measure of performance, percentage salary increase could be viewed to a certain extent as a "surrogate" for the employing firm's assessment of performance. However, its limitations as a "surrogate" must be acknowledged since salary increases can be attributed to causes other than performance; for example, labor market pressures, overall financial condition of the firm, economic conditions. As reported in table 1, the mean salary increase for the sample was 54 percent which normally would be considered exceptionally high in the United States or Europe, but it reflects the high rate of inflation that existed in Mexico at the time this study was conducted. Percentage salary increase was marginally correlated (0.10) with overall self-assessed performance.

---

[7] Prior to the actual distribution of the questionnaires, the top management contact in each firm was asked to indicate both an approximate average salary increase and a range for the managers in their firm. The manager's responses were later checked against that information. Using this referent, none of the responses appeared to be an under- or overstatement of the percentage reported.

Additional information was gathered regarding the level of each manager within the firm, the percentage of foreign ownership of the firm, the industry in which the firm operates, and the functional area in which the respondent works.

III. Administration and Application

Subjects

The subjects were 83 managers from the Mexico City area, working in 21 different companies, varying in size, industry, and degree of foreign ownership. The respondents represented eight different functional areas with 34 managers working in the financial area.[8] Six respondents were women. In terms of nationality, 82 were Mexican citizens, and one was a Mexican resident. Table 1 provides additional descriptive information, including aggregate information regarding overall satisfaction, overall performance, and internality from the questionnaires.

Of the 21 companies represented in the study, seven are in services, two are holding companies, nine are in manufacturing, one is a merchandising firm, and one has characteristics of both merchandising and services. One of the holding companies specialized in the transportation of fluids, mostly hydrocarbons, and was entirely Mexican-owned, except for one subsidiary. The other holding company was highly diversified, with varying degrees of foreign ownership (all 49 percent or less). In terms of respondents, 23 percent were in services, 8.4 percent in merchandising, 60 percent in manufacturing, and 8.4 percent in holding companies. The firms also varied in their degree of local ownership: from 100 percent Mexican to entirely foreign-owned.[9] The parent companies of the foreign-owned firms were mostly from the United States, one French, and one Swiss. The study was conducted in 1986, at a time when Mexico was experiencing economic problems.[10]

Administration

The participating companies were initially selected through their relationship with three Mexico City banks.[11] Within each firm, higher-level managers designated the individual managers for participation in this study. The questionnaires were personally distributed by one of the authors to the 86 managers involved in the study who kept them to complete on their own time. The completed questionnaires were also collected by the same author. Personal contact with both the company and the respondents (and follow-up on all instruments) led to a 100 percent response rate. However, of the 86 questionnaires, three were excluded from the study for incomplete responses.

---

[8] This result was probably due to the bank contacts used and to their interaction with the financial area of the firms.

[9] Most foreign-owned firms that are less than 100 percent are officially 49 percent. However, the official 49 percent may understate the actual percentage of foreign ownership.

[10] These economic problems occurred in the aftermath of the 1982 major devaluation of the Peso. There had been a decline in world oil prices (one of Mexico's major source of income from exports), high inflation (>50 percent), and problems related to the servicing of its $98 billion foreign debt.

[11] Although the sample was not randomly drawn, the companies included in this study are representative of a cross-section of Mexican industry.

Application

The following models were used to test the hypotheses. For consistency, two models are identical to those used by Brownell (1982b).

For performance:

$$Y_1=\beta_1+\beta_2X+\beta_3Z+\beta_4|(X-Z)| \tag{1}$$

$$Y_2=\beta_1+\beta_2X+\beta_3Z+\beta_4|(X-Z)| \tag{2}$$

For satisfaction:

$$Y_3=\beta_1+\beta_2X+\beta_3Z+\beta_4|(X-Z)| \tag{3}$$

where:

Y_1 represents performance as measured by the global rating on the Mahoney measure,
Y_2 represents performance as measured by the percentage salary increase received in the past year,
Y_3 represents satisfaction as measured by the overall measure from the Minnesota Satisfaction Questionnaire (MSQ),
X represents a standardized locus of control score $[(X_i-\overline{X})/\sigma_X]$,
Z represents standardized budgetary participation $[(Z_i-\overline{Z})/\sigma_Z]$, and
$|(X-Z)|$ represents the interaction between locus of control and participation measured as the absolute value of the difference between X and Z.

These models test the effect of participation, locus of control, and their interaction on performance and satisfaction. The interaction term matches the internality measure and the characteristics of the firm's budgetary process. This interaction term is preferred because of *a priori* expectations associated with the manner in which certain locus of control and participation combinations impact satisfaction and performance[12] (Brownell 1982b). For example, when high scores on the locus of control internality dimension (internals) are associated with low scores on the participation measure (high participation), large absolute differences result. This also occurs when low scores on the locus of control internality dimension (externals) are associated with high scores on the participation instrument (low participation). Both of these combinations are expected to lead to higher satisfaction and to higher performance (hypotheses Ho3 and Ho4). Externals are expected to perform better and to experience higher levels of satisfaction under a budgeting system that requires a low level of participation. Internals are expected to perform better and to be more satisfied under a more participative budgetary process. Lower satisfaction and poorer performance would be expected when there is a conflict between a manager's locus of control characteristic and the participative style of the firm.

---

[12] This explanation of the interaction term is similar to Brownell's (1982b) and is based on matching the personality trait with the level of participation. The more frequently used multiplicative interaction term does not provide a good measure of this matching condition. While it may work for extreme values of X and Z, the multiplicative interaction term does not perform well across the entire range.

Table 2
Regression Results for Performance and Satisfaction

Dependent Variable	β_1	β_2	β_3	β_4	R^2	*df*	F-value
Overall Performance							
Coefficient	6.592	−0.104	−0.262	0.408	0.104	73	2.820
Standard Error		0.125	0.133	0.156	0.067*		
Probability		ns	0.052	0.011			
Percentage Salary Increase							
Coefficient	49.504	−0.158	−12.133	3.755	0.241	58	6.131
Standard Error		2.576	2.835	3.266	0.202*		
Probability		ns	0.0001	ns			
Overall Satisfaction							
Coefficient	79.761	1.572	−2.831	−0.360	0.122	72	3.344
Standard Error		1.055	1.107	1.306	0.086*		
Probability		ns	0.013	ns			

Note: β_1 = intercept,
β_2 = coefficient for standardized Internality,
β_3 = coefficient for standardized Participation, and
β_4 = coefficient for the Interaction term.
* Adjusted R^2

IV. Results

The results of the regression analysis are reported in table 2. All three regressions are statistically significant ($p < 0.05$).[13]

Overall Performance

The interaction coefficient is significant at the 0.01 level and the participation coefficient is significant at the 0.05 level.[14] These results indicate that overall performance is associated with both locus of control and participation. Further, it seems that the negative effect of a poor match for externals (high participation) can be partially mitigated by the positive effect of participation on overall performance. These results are consistent with Brownell (1981, 1982b). In addition, both the coefficient of participation and the coefficient of the interaction term are in the expected direction. The negative participation coefficient is expected because Milani's measure gives a high score for low participation and a low score for high participation. As indicated earlier, the interaction coefficient was expected to be positive. Based on these results, both the null hypotheses, Ho1 and Ho3, are rejected for overall performance. This finding suggests that, for overall performance and when all levels are combined, no culture effect is present.[15]

---

[13] The power of the test can only be estimated since it relies on the "true" population parameters, which are unknown. The power of the test was estimated for the overall regressions, given R^2, the sample size, and the number of parameters included in the model, utilizing a method described by Cohen (1977, 414–415). The estimates of the power of the regressions are 69 percent for overall performance, 96 percent for percentage salary increase, and 76 percent for overall satisfaction.

[14] As a statistical check on its validity and reliability, the Mahoney et al. (1963) performance measure was tested by regressing the overall performance rating on the eight subcomponents of performance, a method used by Brownell (1982b) and Brownell and McInnes (1986). The eight dimensions should explain at least 55 percent of overall performance (Mahoney et al. 1963). This study's results indicate that the subcomponents of performance accounted for 65 percent of the variance in overall performance.

Percentage Salary Increase

In this regression, the coefficients are of the expected sign, but only the coefficient for the participation term is significant. Therefore, hypothesis Ho1 is rejected for percentage salary increase as an external measure of performance, but hypothesis Ho3 cannot be rejected. This result is consistent with the hypothesis that participation is significantly related to salary increases. The locus of control did not appear to affect salary increases significantly. Although the results for this external performance measure are not fully consistent with the previous results reported for overall performance, they are not contradictory (since all the estimated coefficients had the same sign).

Overall Satisfaction

Three extreme outliers which had strongly biased answers to the MSQ questionnaire were excluded from the analysis. Also in this regression, the participation coefficient is statistically significant. The interaction term is not significant and fails to show the expected sign. These results lead to a rejection of hypothesis Ho2 (but not of hypothesis Ho4) and to the conclusion that participation alone positively affects the overall satisfaction of the Mexican managers.[16] It seems that internals are not significantly more satisfied with high budgetary participation. This result differs from Brownell's (1982b)[17] and may be attributed to the difference in culture as discussed later in this paper.

Tests for Hypotheses Ho5–Ho8

Additional tests were performed by dividing the sample into two groups based on the nature of the effect being analyzed. After performing regression analysis individually on each group using the previously discussed models, the two regression lines were statistically compared using a formal test to determine the significance of the difference. Each effect is discussed separately.

---

[15] The regression models were also used to test the same relationships with the eight subdimensions of performance. In accordance with Brownell's (1982b) study the results are not consistent for all dimensions. This lack of consistency is attributed to the fact that each dimension measures a different aspect of performance. The regression is significant in explaining the variation for four dimensions: planning, investigating, staffing, and representing. The evaluating dimension regression exhibits a marginal level of significance (0.1052). However, none of these regressions show both the participation and the interaction term as significant. For the dimensions of planning, evaluating, staffing, and representing, participation is significant at 0.05 level but the interaction term is not significant. For investigating, only the interaction term is significant. Therefore, the results for the subsets of performance do not entirely confirm the results for overall performance. As a group, these results suggest that participation alone dominates as an explanatory variable. This finding is similar to that of Brownell (1982b).

[16] Tests were also performed with two subsets of satisfaction: intrinsic and extrinsic satisfaction. Both regression models are significant at the 0.10 level and only participation is significant in both regressions. This finding effectively confirms the results reported for the overall satisfaction measure.

[17] The power of the test becomes an important issue when reporting nonsignificant results, especially when contrasting a study's results with those of other researchers. The power of the *t*-test of the individual beta coefficients was estimated using Owen's (1962) chart of the power function of the *t*-test and Brownell's (1982b) beta coefficient for the interaction term as an estimator of the population beta. The power level estimate for the interaction coefficient is 70 percent for satisfaction.

Table 3

Regression Results for Performance and Satisfaction by Levels

Dependent Variable	β_1	β_2	β_3	β_4	R^2	*df*	*F-value*
Overall Performance							
When Level = 1							
Coefficient	6.643	−0.186	−0.146	0.625	0.298	25	3.543
Standard Error		0.176	0.160	0.197	0.214*		
Probability		ns	ns	0.004			
When Level = 2, 3							
Coefficient	6.557	−0.026	−0.371	0.245	0.084	42	1.278
Standard Error		0.165	0.195	0.221	0.018*		
Probability		ns	0.065	ns			
Percentage Salary Increase							
When Level = 1							
Coefficient	49.557	−4.330	−16.903	3.142	0.470	20	5.910
Standard Error		4.778	4.184	5.270	0.390*		
Probability		ns	0.001	ns			
When Level = 2, 3							
Coefficient	49.012	1.585	−7.464	4.233	0.109	34	1.381
Standard Error		3.048	3.869	4.141	0.030*		
Probability		ns	0.062	ns			
Overall Satisfaction							
When Level = 1							
Coefficient	78.455	0.949	−4.724	−2.079	0.336	24	4.056
Standard Error		1.863	1.677	2.070	0.254*		
Probability		ns	0.010	ns			
When Level = 2, 3							
Coefficient	80.007	1.677	−1.729	0.951	0.073	41	1.082
Standard Error		1.242	1.446	1.629	0.006*		
Probability		ns	ns	ns			

Note: Level # = the hierarchical position of the manager in the firm, relative to top management,
β_1 = intercept,
β_2 = coefficient for standardized Internality,
β_3 = coefficient for standardized Participation, and
β_4 = coefficient for the Interaction term.
* Adjusted R^2

Levels

Managers were categorized in one of three distinct levels. Levels were measured on the questionnaire by the distance between the highest level of decision-making in the organization and the levels of the managers themselves. Level 1 indicated that the subject reported directly to the top manager in the organization, level 2 indicated that the subject was two levels below top management, and so forth. For the purposes of this analysis, the sample was divided into two groups with level-one managers in one group and level-two and level-three managers in the second group.

As reflected in table 3, only the regression results for level-one managers are statistically significant for all three dependent variables. In addition, the regressions for overall performance and overall satisfaction produced different results concerning the influence of participation and the interaction term. The Chow test indicated that the regression estimates were statistically different for both overall performance at the 0.05 level and overall satisfaction at the 0.01 level leading to the rejection of hypothesis Ho5 for these two dependent variables.

For higher-level managers, the interaction term was significant in explaining overall performance. By contrast, participation was significant in explaining overall performance for lower-level managers. However, the direction of the effect was the same in each level, implying that the Mexican manager's level may influence the strength of the effect of both participation and the interaction term in explaining overall performance. The results for the higher-level Mexican managers were similar to Brownell's results (1982b) for the U.S. managers who appeared to be second-level managers.

Only the regression for higher-level managers confirmed fully the prior results for overall satisfaction. For percentage salary increase, both levels fully confirmed the prior results. In addition, the two regressions did not statistically differ, and hypothesis Ho5 cannot be rejected for percentage salary increase.

These findings suggest a managerial level effect for overall performance and for overall satisfaction. The results for level-one managers are consistent with the hypothesized effect of "matching" locus of control and level of participation. It seems that higher-level external managers perform better under low participation situations while the reverse holds true for their internal counterparts at the same level.

Thus, when the managerial level is considered, cultural differences cannot be rejected. This finding is in accordance with Hofstede's framework, in which it was expected that some Mexican managers would prefer a nonparticipative budgetary structure. Consequently, Brownell's model may be more descriptive of higher than of lower managerial levels.

Percentage of Foreign Ownership

The data for the Mexican managers was divided into two subsets based on whether the employing company was entirely foreign-owned or not (less than 100 percent foreign-owned capital). Table 4 presents the results for the two regressions involving each dependent variable.

For the entirely foreign-owned group, the three regressions are not significant, but the prior results are confirmed for the other group. In addition, the Chow test indicates that the two groups differ significantly for all three regressions (0.05 for overall performance, 0.01 for percentage salary increase, and 0.01 for overall satisfaction). Therefore, the null hypothesis, Ho6, is rejected. This finding leads to the conclusion that there may be, as hypothesized, cultural effects in foreign-owned firms due to the cultural interface existing in some of their subsidiaries.

Tests for the Effect of Other Variables

Similar tests were employed on two industry groups, one group consisting of all subjects employed by manufacturing firms and the second group comprising all subjects employed by nonmanufacturing firms (services and merchandising). No statistical difference was present between the regressions; therefore, the null hypothesis, Ho7, cannot be rejected. There is no apparent industry effect.

Table 4
Regression Results for Performance and Satisfaction by Percent Foreign Ownership

Dependent Variable	β_1	β_2	β_3	β_4	R^2	*df*	*F-value*
Overall Performance							
When % < 100							
Coefficient	6.441	−0.005	−0.282	0.569	0.178	42	3.022
Standard Error		0.143	0.156	0.193	0.119*		
Probability		ns	0.078	0.005			
When % = 100							
Coefficient	6.879	−0.102	−0.393	−0.006	0.070	20	0.504
Standard Error		0.374	0.351	0.428	−0.069*		
Probability		ns	ns	ns			
Percentage Salary Increase							
When % < 100							
Coefficient	51.089	3.341	−14.101	4.586	0.337	34	5.749
Standard Error		3.122	3.479	4.310	0.278*		
Probability		ns	0.0003	ns			
When % = 100							
Coefficient	51.034	−12.915	−11.821	6.418	0.313	14	2.130
Standard Error		6.418	6.083	7.002	0.166*		
Probability		0.064	0.072	ns			
Overall Satisfaction							
When % < 100							
Coefficient	75.412	1.815	−5.218	2.459	0.304	42	6.122
Standard Error		1.169	1.249	1.537	0.255*		
Probability		0.128	0.0001	0.117			
When % = 100							
Coefficient	82.779	2.816	0.945	−1.407	0.054	19	0.359
Standard Error		2.727	2.582	3.118	−0.096*		
Probability		ns	ns	ns			

Note: % = percentage of foreign ownership of the firm,
β_1 = intercept,
β_2 = coefficient for standardized Internality,
β_3 = coefficient for standardized Participation, and
β_4 = coefficient for the Interaction term.
* Adjusted R^2

Additional tests were performed with accounting and finance managers in one group and with all other functional areas managers in another. There was no evidence of a statistically significance department effect; therefore, the null hypothesis, Ho8, is not rejected.

The lack of either an industry or a functional area effect increases confidence that the findings discussed previously (hypotheses Ho1–Ho6) are not limited to a particular industry group or to a particular functional area.

V. Discussion

This study provides evidence that cultural differences may lead to different responses to participative budgeting and that various groups within a culture may react differently to a particular budgeting system. Managerial level and the percentage of foreign ownership were useful in delineating the specific settings and types of managers (internals vs. externals, higher vs. lower levels) that proved to be most effective for participation in the budgeting process. Performance for Mexican managers employed by firms owned or controlled by Mexican interests was increased by participation in the budgeting process. At higher levels, internal managers appear to be more positively affected by budgetary participation than are their external counterparts. However, these firms may not find that increased performance will be matched by increased job satisfaction for all managers; this last effect pertains to high-level managers only.

Based on this study's results, firms owned by outside interests (e.g., U.S. or European) may wish to exercise caution when applying participative budgeting techniques in Mexico, perhaps involving upper-level managers in the implementation process. Davis (1969) indicates that when a management technique (such as participative budgeting) is exported to a new cultural environment, implementation is more successful when introduced by a local rather than a foreign manager. Input from local managers as to the most appropriate way of introducing the technique, as well as their suggestions in adapting it to improve its acceptance, should be sought. Local managers who all share the same organizational "reality" with subordinates are more likely to effectively tailor the technique to that environment.

A limitation of our study is that we did not attempt to assess the causes of the different results observed for foreign-owned companies. Future research in this area might specifically address this issue as well as determining how a system could be designed to meet the needs of a particular environment while avoiding the assumption that a particular model must serve as a standard. Based on a review of both cross-cultural management and cross-cultural budgeting literature, it would seem necessary to conduct additional exploratory research to provide a basis for such development. Other limitations of the study are those traditionally associated with field research including the nonrandom selection of subjects and companies and its effect on the generalizability of the study.

References

Adler, N. J. 1983. Cross-cultural management: Issues to be faced. *International Studies of Management and Organizations* 13 (No. 1/2): 7–45.

———. 1984. Understanding the ways of understanding: Cross-cultural management methodology reviewed. *Advances in International Comparative Management.* Ed. R. N. Farmer, 31–67. JAI Press.

———, R. Doktor, and S. G. Redding. 1986. From the Atlantic to the Pacific century: Cross-cultural management reviewed. Yearly Review of Management of the *Journal of Management* 12 (No. 2): 295–318.

Argyris, C. 1952. *The Impact of Budgets on People.* The School of Business and Public Administration, Cornell University.

Baker, S., A. G. Bard, and E. Weiner. 1988. The magnet of growth in Mexico's North. *Business Week* 3055 (June) 5: 48–50.

Brownell, P. 1981. Participation in budgeting, locus of control and organizational effectiveness. *The Accounting Review* 56 (October): 844–60.

———. 1982a. Participation in the budgeting process—when it works and when it doesn't. *Journal of Accounting Literature* 1 (Spring): 124–53.

———. 1982b. A field study examination of budgetary participation and locus of control. *The Accounting Review* 57 (October): 766–77.

———. 1983. Leadership style, budgetary participation and managerial behavior. *Accounting, Organizations and Society* 8 (No. 4): 307–21.

———, and M. Hirst. 1986. Reliance on accounting information, budgetary participation, and task uncertainty: Tests of a three-way interaction. *Journal of Accounting Research* 24 (Autumn): 241–49.

———, and M. McInnes. 1986. Budgetary participation, motivation and managerial performance. *The Accounting Review* 61 (October): 587–600.

Bryan, J. F., and E. A. Locke. 1967. Goal setting as a means of increasing motivation. *Journal of Applied Psychology* 51 (June): 274–77.

Chenhall, R. H. 1986. Authoritarianism and participative budgeting: A dyadic analysis. *The Accounting Review* 61 (April): 263–72 .

———, and P. Brownell. 1988. The effect of participative budgeting on job satisfaction and performance: Role ambiguity as an intervening variable. *Accounting, Organizations and Society* 13 (No. 3): 225–33.

Cherrington, D. J., and J. O. Cherrington. 1973. Appropriate reinforcement contingencies in the budgeting process. *Journal of Accounting Research* 11 (Supplement): 225–53.

Child, J. 1981. Culture, contingency and capitalism in the cross-national study of organizations. *Research in Organizational Behavior.* Eds. L. L. Cummings, and B. M. Shaw, 303–56. JAI Press.

———, and M. Tayeb. 1982–83. Theoretical perspectives in cross-national organizational research. *International Studies of Management and Organizations* 12 (No. 4): 23–70.

Cohen, J. 1977. *Statistical Power Analysis for the Behavioral Sciences.* Academic Press.

Collins, F. P. 1978. The interaction of budget characteristics and personality variables with budgetary responses attitudes. *The Accounting Review.* 53 (April): 324–35.

Daley, L., J. Jiambalvo, G. L. Sundem, and Y. Kondo. 1985. Attitudes toward financial control systems in the United States and Japan, *Journal of International Business Studies* 16 (Fall): 91–110.

Davis, S. M. 1969. U.S. versus Latin America: Business and culture. *Harvard Business Review* 47 (November–December): 88–98.

Diaz-Guerrero, R. 1975. *Psychology of the Mexican: Culture and Personality.* U. T. Press.

———. 1984. Psychology for the Mexican or for the masses. *Psychology, A Quarterly Journal of Human Behavior* 21 (No. 2): 1–7.

Dunk, A. S. 1989. Budget emphasis, budgetary participation and managerial performance: A note. *Accounting, Organizations and Society* 14 (No. 4): 321–24.

Garnier, G. 1984. The autonomy of foreign subsidiaries: Environmental and national influences. *Journal of General Management* 10 (Autumn): 57–82.

Gilbreath, K. 1986. A businessman's guide to the Mexican economy. *Columbia Journal of World Business* 21 (Summer): 3–14.

Goodstadt, B. E., and L. A. Hjelle. 1973. Power to the powerless: Locus of control and the use of power. *Journal of Personality and Social Psychology* 27 (August): 190–96.

Gurin, P., G. Gurin, R. C. Lao, and M. Beattie. 1969. Internal-external control in the motivational dynamics of Negro youth. *Journal of Social Issues* 25 (No. 3): 29–53.

Heneman, H. G. III. 1974. Comparisons of self and superior ratings of managerial performance. *Journal of Applied Psychology* 59 (No. 5): 638–42.

Hofstede, G. H. 1967. *The Game of Budget Control.* Van Gorum.

———. 1980. *Culture's Consequences: International Differences in Work-Related Values.* Sage Publications.

———. 1983. National cultures in four dimensions. *International Studies of Management and Organization* 13 (No. 1/2): 46–74.

———. 1987. The cultural context of accounting. *Accounting and Culture.* Ed. B. E. Cushing, 1–11. American Accounting Association.

Hui, C. H. 1982. Locus of control: A review of cross-cultural research. *International Journal of Intercultural Relations* 6 (No. 3): 301–23.

———, and H. C. Triandis. 1983. Multistrategy approach to cross-cultural research: The case of locus of control. *Journal of Cross-Cultural Psychology* 14 (March): 65–83.

———, and ———. 1985. Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology* 16 (June): 131–52.

Inzerelli, G., and A. Laurent. 1983. Managerial views of organization structure in France and the U.S.A. *International Studies of Management and Organizations* 13 (No. 1/2): 97–118.

Kanungo, R. N., and R. W. Wright. 1981. *A Cross-Cultural Comparative Study of Managerial Job Attitudes.* ASAC Conference.

Kenis, I. 1979. Effects of budgetary goal characteristics on managerial attitudes and performance. *The Accounting Review* 54 (October): 707–21.

La Rosa, J. 1986. Escala de Locus de Control y Autoconcepto: Construccion y Validacion. Tesis de Doctorado. Universidad Nacional Autonoma de Mexico.

Laurent, A. 1983. The cultural diversity of Western conceptions of management. *International Studies of Management and Organizations* 13 (No. 1/2): 75–96.

Lefcourt, H. M. 1976. *Locus of Control: Current Trends in Theory and Research.* Wiley.

———. 1982. *Locus of Control: Current Trends in Theory and Research*; Second Edition. Lawrence Erlbaum and Associates.

Levitt, T. 1983. The globalization of markets. *Harvard Business Review* 61 (May–June): 92–102.

Licata, M. P., R. H. Strawser, and R. B. Welker. 1986. A note on participation in budgeting and locus of control. *The Accounting Review* 61 (January): 112–17.

Locke, E. A., and D. M. Schweiger. 1979. Participation in decision making: One more look. *Research in Organizational Behavior.* Ed. B. M. Staw, 265–339. JAI Press.

Mahoney, T. A., T. H. Jerdee, and S. J. Carroll. 1963. *Development of Managerial Performance: A Research Approach.* Southwestern Publishing.

Merchant, K. A. 1981. The design of the corporate budgeting system: Influences on managerial behavior and performance. *The Accounting Review* 56 (October): 813–29.

Mia, L. 1988. Managerial attitude, motivation and the effectiveness of budget participation. *Accounting, Organizations and Society* 13 (No. 5): 465–75.

———. 1989. The impact of participation in budgeting and job difficulty on managerial performance and work motivation: A research note. *Accounting, Organizations and Society* 14 (No. 4): 347–57.

Milani, K. 1975. The relationship of participation in budget-setting to industrial supervisor performance and attitudes: A field study. *The Accounting Review* 50 (April): 274–84.

Miner, J. B. 1968. Management appraisal: A capsule review and current references. *Business Horizons* 11 (October): 83–96.

Mitchell, T. R., C. M. Smyser, and S. E. Weed. 1975. Locus of control: Supervision and work satisfaction. *Academy of Management Journal* 18 (September): 623–31.

Owen, D. B. 1962. *Handbook of Statistical Tables.* Addison-Wesley Publishing.

Parker, J. W., E. K. Taylor, R. S. Barrett, and L. Martens. 1959. Rating scale content: III. Relationships between supervisory- and self-ratings. *Personnel Psychology* 12 (Spring): 49–63.

Phares, E. J. 1976. *Locus of Control in Personality.* General Learning Press.

Prien, E. P., and R. E. Liske. 1962. Assessments of high-level personnel: III. A comparative analysis of supervisor ratings and incumbent self-ratings of job performance. *Personnel Psychology* 13 (Summer): 187–94.

Riding, A. 1985. *Distant Neighbors: A Portrait of the Mexicans.* Alfred Knopf.

Rotter, J. B. 1966. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs* 80 (No. 1): 1–28.

Runyon, K. E. 1973. Some interactions between personality variables and management styles. *Journal of Applied Psychology* 57 (June): 288–94.

Stening, B. W. 1979. Problems in cross-cultural contact: A literature review. *International Journal of Intercultural Relations* 3 (No. 3): 269–313.

Swieringa, R. J., and R. H. Moncur. 1975. *Some Effects of Participative Budgeting on Managerial Behavior.* National Association of Accountants.

Thornton, G. C. 1968. The relationship between supervisory- and self-appraisals of executive performance. *Personnel Psychology* 21 (Winter): 441–56.

Triandis, H. C. 1959. Cognitive similarity and interpersonal communication in industry. *Journal of Applied Psychology* 43 (October): 321–26.

———. 1960. Cognitive similarity and communication in a dyad. *Human Relations* 13 (May): 175–83.

———. 1972. *The Analysis of Subjective Culture*. Wiley.

Weiss, D. J., R. V. Davis, G. W. England, and L. H. Lofquist. 1967. *Manual for the Minnesota Satisfaction Questionnaire*. Minnesota Studies in Vocational Rehabilitation.

Weisz, J. R., F. M. Rothbaum, and T. C. Blackburn. 1984. Standing out and standing in: The psychology of control in America and Japan. *American Psychologist* 39 (September): 955–69.

Williams, W. E., and D. A. Seiler. 1973. Relationship between measures of effort and job performance. *Journal of Applied Psychology* 57 (No. 1): 49–54.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 100–119

Configural Information Processing in Auditing: The Role of Domain-Specific Knowledge

Clifton E. Brown
Ira Solomon
University of Illinois at Urbana-Champaign

SYNOPSIS: A prominent stream of audit judgment research employs the policy-capturing paradigm and uses analysis of variance (ANOVA) to construct mathematical models of individual auditors' evaluations of internal control. Based on these "as if" representations, it may appear that auditors evaluate control systems by independently processing individual controls as opposed to the pattern (or configuration) of control features. Indeed, with a few noteworthy exceptions, virtually no evidence has been reported of auditors configurally processing information while evaluating controls or assessing control risk. One such exception is a recent paper by Brown and Solomon (1990). Therein, it was suggested that judicious usage of context- and task-specific knowledge is necessary to guide the development of experiments capable of detecting configural information processing.

The present study generalizes the concept that auditors will exhibit configural information processing in situations in which domain-specific knowledge implies that it is appropriate. Two experiments using auditor-subjects and investigating configural information processing are described. The task of the first experiment is assessment of account misstatement risk given evidence produced by specified substantive audit procedures; and the task of the second experiment is such assessment given the results of specified analytical audit planning procedures. Domain-specific knowledge suggests configural information processing strategies that will be manifest (within ANOVA) as a compensatory-form ordinal interaction for the former

This research was funded by KPMG Peat Marwick through a Research Opportunities in Auditing grant. We wish to acknowledge the comments of Mike Shields, participants in a research colloquium at the University of Iowa, and of our colleagues Jon Davis, John Hansa, Lisa Koonce, and Fred Neumann. Last, we note that without the assistance of KPMG Peat Marwick (especially John Naughton and the Chicago office), this study would not have been possible.

Submitted April 1989.
Revisions received March and August 1990.
Accepted August 1990.

task and as a disordinal interaction for the latter task. Experimental results are consistent with expectations and are especially striking for the second experiment. Twenty of the second experiment's 22 auditor-subjects apparently used the predicted configural strategy, and ANOVA models of these auditors' risk assessments attributed on average about 40 percent of total judgment variation to the single predicted interaction of analytical procedure results (trial balance figure comparisons).

An implication of these results, in concert with Brown and Solomon (1990), is that auditor information processing is more complex than has been reported by prior studies. That is, across a wide variety of tasks, many auditors apparently are able to configurally process information. The results also imply considerable between-task and between-auditor configural processing differences (magnitude and prevalence), which anecdotal evidence suggests is due to domain-specific factors. Two factors, the rate-of-substitution between information cues (substitutable audit procedures or controls) and when auditors learn the domain-specific knowledge requisite for configural information processing, seem particularly important and could be worthwhile foci for future research.

Key Words: *Configurality, Information processing, Audit judgment, Domain-specific knowledge.*

Data Availability: *Data analyzed in this study may be obtained on request from the authors.*

RESEARCHERS have recognized the essential nature of domain-specific knowledge for understanding judgment and decision-making strategies (Newell and Simon 1972; Einhorn and Hogarth 1981; Payne 1982; Frederick and Libby 1986). Recently, Brown and Solomon (1990) argued that domain-specific knowledge is crucial for developing expectations for, and thus demonstrating, configural information processing. That is, both context- and task-specific knowledge need to be used in determining the judgment tasks that are more effectively and efficiently accomplished employing configural information processing. Subsequently, the researcher needs to construct experimental tasks in which expert-subjects will be motivated (able) to process information in a configural fashion. Implicit in this reasoning is the assumption that if context- and task-specific knowledge implies that configural information processing is appropriate, judges who possess such knowledge will have learned such configurality. Using knowledge specific to control risk assessment in the task of audit planning, Brown and Solomon (1990) developed expectations for two specific forms of configural information processing and reported experimental evidence of such configural strategies.

The present paper generalizes the concept that auditors will exhibit configural information processing in situations in which domain-specific knowledge implies that such configurality is appropriate. In the next section, we review prior configural processing research and develop the importance of domain-specific knowledge for this type of research. In the third and fourth sections, we describe two experiments: each begins with a discussion of auditing knowledge expected to be used by subjects when performing the associated audit judgment task and is followed by the hypotheses to be tested, a description of the experiment and a presentation of experimental results. In the final section, we discuss the results of the two experiments (both individually and jointly) and describe implications for future research.

I. Prior Research

With few exceptions, prior policy-capturing psychology studies have *not* produced evidence that individuals, either experts or novices, configurally process available information (Hammond and Summers 1965; Goldberg 1968; and Slovic et al. 1977).[1] The series of papers on auditors' evaluation of internal controls is the most distinct example of auditing studies that have investigated configural information processing, at least as one of several related research issues. Auditing researchers, working in the psychology paradigm of policy-capturing, employed ANOVA to model auditor-subjects' judgments and searched for evidence of configural cue usage by examining the statistical significance of the judgment variation attributable to interaction terms in such models (Ashton 1982, 1983; Libby 1981).

Although the first such study was conducted by Ashton (1974), many extensions have been reported (Ashton and Kramer 1980; Ashton and Brown 1980; Hamilton and Wright 1982; Hall et al. 1982; Trotman et al. 1983). All of these studies essentially were unable to produce evidence that auditors configurally processed available information. More specifically, the results reported in Ashton (1974) are typical: whereas 60.3 percent of the auditor-models attributed a statistically significant portion of judgment variation to one or more interactions, the *single* interaction most frequently found to be significant was so in only 23.8 percent of the auditor-models. Further, although across auditor-models the mean judgment variation attributable to the sum of *all* interactions was 6.4 percent, the highest mean judgment variation attributable to any *single* interaction was less than 4 percent when averaged over those auditor-models in which it was significant (less than 1 percent when averaged over all auditor-models).

Hoffman (1968, 63) observed that for a variety of reasons, "an undirected search for configural relationships in a finite set of data is fraught with difficulties and virtually doomed to disappointment." As the complexity of a decision context increases (e.g., the number of dimensions or terms increases), the problem of parameter estimation also increases to the point that interactions, without a priori expectations, can become very difficult to detect. Brown and Solomon (1990) speculated that a major reason for the inability of prior auditing studies to detect configural information processing is that they did not provide a priori expectations for specific cue interactions. Examining the procedures and materials employed in prior auditing studies, it was observed that none of the studies captured two or more control features that related to the same control objective in such a way that configural information processing would be a salient, appropriate strategy.[2] Consequently, there is little reason for an auditor to have done anything other than what was observed in prior studies: evaluate independently the effect of each control.

---

[1] While policy-capturing studies generally have failed to detect configural information processing, there are a few exceptions. For example, configurality has been observed in stockbrokers' judgments of stock growth potential (Slovic 1969), psychiatric medical professionals' judgments of granting weekend passes to patients (Rorer et al. 1967), and individuals' judgments of moral behavior (Leon et al. 1973). Outsde of the policy-capturing paradigm, Einhorn et al. (1979) have noted that process-tracing models suggest that information search often is configural. In addition, studies of multiple cue probability learning (MCPL) have reported that individuals can learn to configurally process information, at least in part, through repeated sequences of judgments followed by criterion value feedback (Castellan 1977; Camerer 1987). MCPL studies, however, have not used an existing body of professional knowledge to identify the specific form and nature of the configurality expected to be learned by subjects. Further, we do not consider the MCPL approach to be appropriate when studying judgments in professional contexts such as auditing because professional judges generally have an existing level of expertise and criterion values rarely are observable (at least in a timely fashion).

Brown and Solomon (1990) argued that domain-specific knowledge has an important role in developing a priori expectations for configural information processing. Indeed, domain-specific knowledge has been discussed in a number of recent studies in which the amount and organization of such knowledge are identified as the major components of expertise in disciplines such as accounting and auditing (Frederick and Libby 1986; Davis and Solomon 1989). While such discussions usually are in general terms, for our purposes, domain-specific knowledge must be considered at a more fundamental level. That is, a condition necessary for the researcher to observe configural information processing is that extant domain-specific knowledge must not only imply that configural information processing is appropriate, but it also must state the form and nature of such configurality.

Brown and Solomon (1990) created a control risk assessment task for which expectations were developed, based on context- and task-specific knowledge, about two interactions involving cash disbursements controls. In particular, using the auditing concepts of compensating and amplifying controls, they predicted that either a preventive and detective control (or both) would interact with a separation-of-duties control. Further, they predicted that such interaction(s) would be ordinal in either compensatory- or amplifying-form (defined below), depending upon whether the subjects perceived such controls as primarily compensating a separation-of-duties weakness or amplifying a separation-of-duties strength. The experimental results were consistent with usage of such configural information processing strategies by many of the auditor-subjects. However, while 40.5 percent of the auditors' judgment-models attributed a significant portion of judgment variation to one or both of the two *expected* interactions (in the forms expected), the mean judgment variation attributable to the two specific interactions was modest (8.16 percent).

II. Experiment One

The context of the audit task employed in experiment one is evaluation of evidence produced by substantive audit procedures (specifically, assessment of account misstatement risk given a partially completed portion of an accounts receivable audit program). The auditing knowledge required to appropriately perform this task implies an ordinal interaction of a particular form.

Knowledge Required for Evaluation of Substantive Audit Procedures

Consider a situation in which an auditor is concerned with the risk that a particular account (accounts receivable in the present experiment) is presented in conformance with generally accepted accounting principles. As reflected in the audit risk model (AICPA 1983), both indirect evidence (i.e., evidence relevant to inherent and control risk assessments) and direct evidence obtained from substantive procedures (i.e., tests of details and analytical procedures) should be collected and evaluated to appraise the risk of erroneous financial statement presentation. Prior research (and practice) suggests that the circumstances of specific audit engagements, rather than general rules of thumb should dictate which audit procedures are appropriate as well as when and to what extent they should be performed (Mock and Wright 1982).

---

[2] Ashton (1974) is representative of the prior audit studies. Two separation-of-duties and an internal audit control, potentially interrelated in an evaluation of payroll "internal control strength," were included in that study. The internal audit control, however, was not explicitly related to payroll control objectives, and the control objective itself was a global assessment rather than an asssessment of a particular audit assertion (such as the payroll checks being issued for unauthorized and/or invalid purposes).

Such a contingency approach to audit program planning may be illustrated as follows (Wright and Mock 1985). The auditor first identifies specific assertions that are to be the focus of attention. If, for example, the auditor's primary concern were the existence of recorded accounts receivable, a useful procedure would be direct confirmation of recorded balances with creditors. While a number of options exist in this regard, positive-format confirmation requests may be sent and, if returned without exceptions noted, a conclusion about the existence of the accounts receivable may be formulated. On the other hand, if a sufficient number of confirmations were not returned or returned with exceptions noted, evidence would have to be obtained by performing other procedures.

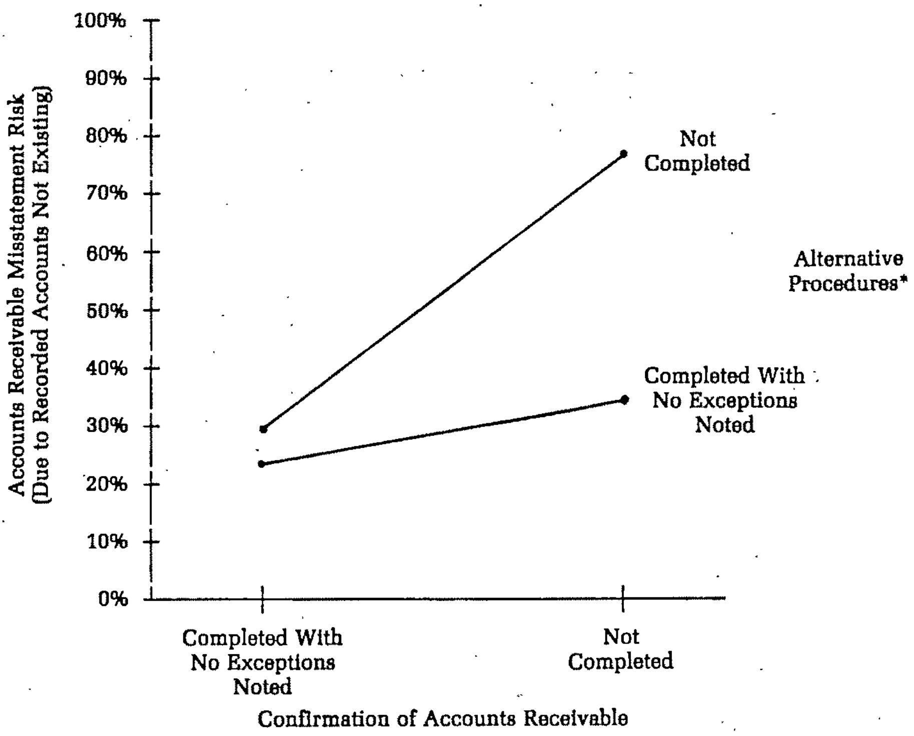
Confirmation therefore may be viewed as a primary procedure, while verification of sales transactions and observation of subsequent collections may be viewed as alternative (or "fall-back") procedures. However, when the debtor is not likely to return the confirmation (e.g., governmental agencies) or is unable to determine reliably the amount owed (e.g., patients within health-care institutions), procedures such as verification of sales transactions and observation of subsequent collections may be viewed as primary procedures. Thus, with respect to the existence assertion and depending upon specific audit circumstances, these audit procedures (confirmations versus verification/observation) may be viewed as substitutable for each other. In addition, because either procedure alone can provide sufficient, competent evidence with respect to "existence," given that one procedure has been performed, there would be little incremental benefit derived from performing the other procedure. However, if neither procedure were performed, the "existence" risk may be sufficiently high to preclude continuation of the audit.

Assessments of misstatement risk (due to lack of existence) reflecting these domain-specific features will appear as a compensatory-form ordinal interaction in an ANOVA model. Ordinality is defined as a *conditionally monotonic* relation between independent and dependent variables (Dawes 1988). In the present context, this means that positive (negative) evidence provided by an audit procedure implies lower (higher) misstatement risk judgments, irrespective of the evidence from other audit procedures. Further, the term *compensatory-form* means that the one audit procedure's (e.g., verification of sales/observation of subsequent collections) at least partial substitutability for another audit procedure (e.g., confirmation of accounts receivable) is greater than any *amplification* that arises when (if) both procedures were performed.[3] Such a situation is presented graphically in figure 1.

---

[3] This form is contrasted with an *amplifying-form* in which one procedure's amplification of another procedure's effectiveness is greater than its substitutability for the other procedure. See Brown and Solomon (1990) for an example in a control risk context.

Figure 1
Example of a Compensatory-Form Ordinal Interaction in the Context of Experiment One



* Verification of Sales Transactions and Observation of Subsequent Collections

Based on this discussion of auditing knowledge required for evaluation of accounts receivable audit procedures, the following alternative-form hypotheses were developed for experiment one:[4]

H1: The proportion of auditors having an above-criterion (defined below) portion of judgment (misstatement risk) variation attributable to an interaction of substitutable audit procedures (i.e., confirmation of accounts receivable and verification of sales transactions/observation of subsequent collections) will be significantly greater than chance.

---

[4] Although, in addition to explained judgment variation, Brown and Solomon (1990) used judgment model prediction differences to assess the substantive import of small but statistically significant interactions, this method is not employed in the present paper. Judgment prediction is not relevant for experiment one because: (1) the anticipated configural term involves auditors' knowledge of when *not* to perform an audit procedure, and (2) institutional features of the experiment one context constrain audit continuance when *no* audit procedures have been performed (see the discussion of effects on misstatement risk of performing two substitutable audit procedures and of performing no audit procedures). With respect to experiment two, judgment prediction is less relevant because the predicted configural term (representing a disordinal relation) is expected to be, and is, quite large.

H2: For auditors having an above-criterion portion of judgment (misstatement risk) variation attributable to the expected interaction, such interaction will be ordinal with a compensatory-form. This interaction will have the following specific attributes:

a. When neither substitutable audit procedure has been performed, misstatement risk assessments will be significantly larger than when a single substitutable procedure has been performed.
b. When both substitutable audit procedures have been performed, misstatement risk assessments will not be significantly smaller than when only a single substitutable procedure has been performed.

Subjects

Subjects were 49 audit seniors with three to four years of audit experience. The subjects were employed by the same national CPA firm and participated either at one of the firm's offices (12 subjects) or while attending a technical training school (37 subjects).

Research Design

The research design was a one-half fractional replication of the factorial manipulation of five information cues involving audit procedures for accounts receivable, each cue at two levels. An ANOVA model was computed for each auditor-subject's risk assessments. Although each auditor-model estimated all five main effects and ten two-way interactions, the higher-order (three-, four-, and five-way) interactions are aliases of the estimated effects and thus are assumed to be negligible. In addition, because each auditor-model has only one observation per cell, such models are determined fully and no error estimate exists. The variation in judgment attributed to each term in an auditor-model, therefore, was computed by dividing the sum of squares for the term by the total sum of squares for the model. Further, an arbitrary criterion of greater than or equal to four percent of total judgment variation was used to determine significance (i.e., terms to which a model attributes judgment variation less than four percent of total variation are assumed to have been caused by random rather than systematic effects).[5]

Materials

A hypothetical audit case was developed in which configural information processing would be consistent with auditing knowledge specific to evaluation of evidence provided by procedures comprising part of the audit program for accounts receivable. The case included background information, specific information concerning procedures (both presented in appendix A), and specimen stimuli presentations.

The subjects were told that they would be given a series of checklists (completed by an auditor or their staff), each representing a partially completed portion of the accounts receivable audit program, and would be asked to assess, given the audit evidence-to-date, the following misstatement risk (elicited on a 100-point scale where 0 was no risk and 100 was maximum risk):

---

[5] Results of a pilot study ($n=12$) employing a full 2^5 factorial design and similar materials and procedures, indicated that effects ≥ 2 percent of total risk assessment variation were significant when using the higher-order (three-, four-, and five-way) interactions as error estimates.

Exhibit 1
Example of an Accounts Receivable Audit Procedures Checklist from Experiment One

Procedure	*Results*
A. Confirm a 95% reliability sample of year-end A/R (positive form).	Not Completed as of This Date
B. Verify sales transactions for and observe collections of a 95% reliability sample of year-end A/R.	Completed With No Exceptions Noted
C. Reconcile A/R subsidiary ledger to the general ledger control account.	Completed With No Exceptions Noted
D. Verify adequacy of allowance for bad debts using analytical procedures.	Completed With No Exceptions Noted
E. Review aged A/R listing for related party transactions, long-term notes, etc.	Not Completed as of This Date
F. Review Board of Directors' meeting minutes and correspondence files to determine if receivables were factored or discounted.	Not Completed as of This Date

Given the audit evidence as represented by the partially completed audit program segment (above), assess the risk that the year-end accounts receivable balance could be materially misstated as a result of recorded accounts NOT EXISTING.

An example of the accounts receivable audit program checklist is presented in exhibit 1. Confirmation of accounts receivable (procedure A) and verification of sales transactions/observation of subsequent collections (procedure B) were intended to be substitutable accounts receivable audit procedures and thus are the constituent variables of the interaction expected in the first hypothesis. Procedures C through F were intended to be accounts receivable audit procedures less directly related to the existence assertion. Five of the six audit procedures (A, B, C, E, and F in exhibit 1) were factorially manipulated at two levels each ("completed with no exceptions noted" or "not completed as of this date"), and one procedure (D in exhibit 1) was held constant ("completed with no exceptions noted").

Procedures

The laboratory session for experiment one consisted of two sections: training and testing. Both sections were presented on personal computers, and subjects completed the sections at their own pace (the average time was 38.1 and 24.1 minutes per training and testing section, respectively). The training section began with brief instructions on the personal computer and was followed by an audit procedures evaluation case involving inventories. Each subject evaluated four manipulations of the practice inventories case to gain familiarity using the response scale and two decision aids available in the testing section. The decision aids were intended to reduce subject memory load and to control extraneous judgment variation.[6] The training section continued with presentation of the accounts receivable case (see appendix A) and was followed by a blank copy of the audit program checklist and additional instructions.[7] The subjects then responded to a series of questions designed to stimulate prior thought about each item listed on the checklist and its relation to the specific audit objectives for which they were being asked to make risk assessments.[8]

---

[6] The two decision aids were on electronic file and a logical consistency checker. When assessing risk, the subject had access to an electronic file of checklists that he or she had already evaluated. As the subject worked

In the testing section, subjects were presented sequentially with the 16 checklists from one of the half-replications (randomized over subjects). The order of the checklists (i.e., the cue combinations) in each half-replication was randomized for each subject. In addition, the order of the items on the checklist was counter-balanced; one-half of the subjects received one order (the two substitutable audit procedures were listed first, as in exhibit 1) and the other one-half received a second order (the two substitutable audit procedures were listed last). Finally, the subjects responded to a postexperimental questionnaire.

Results

H1: Expected Interaction. Twenty-seven of the 49 auditor-models (55.1 percent) attributed above-criterion judgment variation to the interaction expected in hypothesis H1 (AB in table 1). This frequency is significantly greater than chance (a 95 percent confidence interval for this frequency is 39.2 percent to 71 .0 percent) and thus is consistent with hypothesis H1. The mean judgment variation attributed by the 27 auditor-models to the expected interaction was 12.17 percent (a 95 percent confidence interval for this mean is 10.00 percent to 14.34 percent).[9]

H2: Form and Nature of the Expected Interaction. As indicated above, 27 auditor-models attributed above-criterion judgment variation to the expected interaction. Visual inspection revealed that all 27 interactions were compensatory-form ordinals,

---

through the checklists, the computer reviewed the assessments of dominated audit evidence combinations for logical consistency (e.g., the combination in which all procedures have been completed satisfactorily dominates the combinations in which only some of the procedures have been completed and thus should be assessed misstatement risk that is no greater than such other combinations). When the computer detected an apparent logical inconsistency, it was displayed and the subject had the option of either changing or maintaining his or her assessment of the current checklist. Previous evaluations could not be changed.

[7] Subjects were instructed to ignore the temporal sequence of the checklists and that the checklists would provide a mixture of possible situations. Further, subjects were told that although in actual practice some situations may occur less frequently than others, they should not allow such frequency differences to affect their risk assessments.

[8] For each listed audit procedure, the subject was asked to think about things that could go wrong or misstatements that could go undetected if the procedure were not performed. Each subject also was asked to write these thoughts down and to indicate the procedure's importance for achieving the specified audit objective.

[9] An analysis of these results' sensitivity to the arbitrary criterion (threshold) used in determining significance of judgment variation attributable to the expected interaction indicated the following:

	Judgment Variation Criterion				
Expected Interaction	≥0%	≥2%	≥4%	≥6%	≥8%
Percent of subjects	100.0	61.2	55.1	51.0	46.9
Mean judgment variation	7.0	11.2	12.2	12.8	13.3
Total number	49	30	27	25	23
Number in expected form	44	30	27	25	23

Using the same testing methods described in the paper, both hypotheses H1 and H2 are supported using any criterion in the range of this sensitivity analysis (0 to 8 percent).

Table 1
Summary of Auditor Judgment Models: Experiment One

	Judgment Models Incorporating:				
	Only Above-Criterion Terms			*All Terms*[1]	
Source[2]	*Mean*[3]	*Std. Dev.*	n[4]	*Mean*[3]	*Std. Dev.*
A	44.99	24.42	49	44.99	24.42
B	19.12	11.89	42	16.71	12.53
C	23.75	19.39	34	17.00	19.09
E	7.05	2.71	7	2.00	2.51
F	6.51	1.80	10	2.13	2.52
All Main Effects	80.20	8.31	49	82.82	8.43
AB	12.17	5.49	27	6.99	7.09
AC	6.94	3.93	10	2.03	3.17
AE	7.63	3.87	3	1.34	2.05
AF	4.06	0.08	2	0.83	1.15
BC	7.47	3.45	6	1.40	2.64
BE	4.25	0.00	1	0.77	1.12
BF	5.99	0.00	1	0.73	1.16
CE	5.04	1.26	2	0.73	1.16
CF	5.25	1.07	4	0.86	1.45
EF	5.84	1.18	6	1.48	1.93
All Interactions	11.23	9.18	49	17.18	8.43
Residual	8.58	4.95	49		

[1] The number of auditor-models (n) equals 49.
[2] Sources are keyed to procedures in exhibit 1. The expected interaction is AB.
[3] Judgment variation attributed to the respective source.
[4] The number of auditor-models for which the specified term is above-criterion.

thus supporting the initial part of hypothesis H2 (also see fn. 9). The mean risk assessments for each level of this interaction are presented in table 2. A one-way repeated measures ANOVA indicates that these cell means are significantly different ($F(3,78) = 154.3$; $p < 0.01$). Tukey's test over the levels of this interaction indicates that the mean risk assessment is significantly higher when neither substitutable audit procedure had been performed (NN in table 2) than when any (or both) audit procedures were performed (NY, YN, and YY in table 2). In addition, when both substitutable procedures had been performed (YY in table 2), the mean risk assessments are not significantly smaller than when either single substitutable procedure had been performed (NY and YN in table 2). Together, these results verify the two attributes specified in the latter parts of hypothesis H2 and thus confirm the specific nature expected for the predicted compensatory-ordinal interaction (also see fn. 9).

Case Design Validity Check. The experiment materials were designed such that specific audit procedures should be involved in the expected configural information processing strategy. Configural strategies involving other procedures, while neither explicitly designed into the experiment nor predicted, were not precluded. Nevertheless, the frequency of such resultant interactions should not exceed that of the expected interaction. Thus, a design validity check would be to compare the frequency and mag-

Table 2
Misstatement Risk Assessments: Experiment One

	Levels of the Expected Interaction			
	YY	*YN*	*NY*	*NN*
Mean ($n=27$)	23.57	29.15	34.90	76.43
Standard Deviation	14.28	14.04	12.79	19.91

Note: The expected interaction is confirmation of accounts receivable (A) by verification/observation of subsequent collections (B). The levels are: YY = both audit procedure A and B have been completed and no exceptions noted; YN = A has been completed, but B has not; NY = B has been completed, but A has not; and NN = neither A nor B has been completed.

nitude of the expected interaction to those of other (nonhypothesized) interactions. As previously indicated, 27 of the 49 auditor-models (55.1 percent) attributed above-criterion judgment variation to the expected interaction. Of the nonhypothesized above-criterion interactions, AC (see table 1) had the greatest frequency (10 of the 49 auditor-models). This frequency is significantly smaller than that for the expected interaction ($Z=3.99$; $p<0.01$). Further, the mean judgment variation attributed by the 27 auditor-models to the expected interaction (12.17 percent) is significantly greater than that attributed by the 10 auditor-models to this nonhypothesized interaction (6.94 percent; $t(35)=2.75$; $p<0.01$).

III. Experiment Two

The context of the task employed in experiment two is evaluation of analytical procedure results during preliminary audit planning. In particular, subjects assessed the risk that accounts in the revenue-and-receipts transaction cycle could be materially misstated given a comparison of selected trial balance figures between the prior and current periods.

Interpretation of Financial Statement Account Analytical Procedures

Consider a situation in which an auditor is performing preliminary audit planning at the auditee's year-end and is concerned with an initial assessment of the risk that accounts in a particular transaction cycle (the revenue-and-receipts cycle for the present study) could be materially misstated. A common analytical procedure in this situation is to compare the current year's unaudited trial balance accounts with those from the prior year (audited). Material changes in the accounts could have been caused by misstatement events and/or nonmisstatement events. The latter event is illustrated by changes in economic circumstances experienced by the auditee during the two periods and changes in the auditee's business and/or accounting policies between the two periods. When additional information obtained from the auditee concerning material change(s) indicates that events from the nonmisstatement category (e.g., changes in economic circumstances) are less likely to have been the cause, the judged likelihood of events in the misstatement category should correspondingly increase, and planned audit procedures should reflect that inference.

A potential cue to the cause(s) of material change(s) in trial balance accounts between periods is the pattern of changes for related accounts. If, for example, the cause of a material change in gross sales were to be a nonmisstatement event (e.g., changes in economic circumstances), one would expect net accounts receivable to exhibit a change in the same direction (i.e., knowing a cause implies its effects). However, observing that gross sales and net accounts receivable materially change in the same direction does not rule out an account misstatement event (i.e., observing effects does not imply a cause). Each category of events by itself could have caused observed consistent changes in these accounts, and thus consistent patterns of changes are *not* highly diagnostic with respect to the risk of material account misstatement. In turn, the likelihood of material account misstatement should not be substantially revised on observing consistent changes in these accounts. On the other hand, if gross sales were to change materially, but net accounts receivable were not to change or were to change materially in the opposite direction from the change in gross sales (i.e., inconsistent changes), nonmisstatement events, as a *single* category, would be less likely to have been the cause. Consequently, all other things equal, auditors should assess greater likelihoods for account misstatement on observing inconsistent changes in these accounts.

Misstatement risk assessments reflecting these domain-specific features suggest a disordinal interaction within an ANOVA model, where the meaning of one variable may *reverse* depending upon the value of another variable (i.e., monotonicity is violated). In the context of experiment two, for example, the risk of account misstatement is greater when net accounts receivable has increased materially than when it has decreased materially, given (in both instances) that gross sales has decreased materially, but the reverse is true when gross sales has increased materially. Such a situation is presented graphically in figure 2.

Based on this discussion of auditing knowledge specific to analytical comparisons of financial statement accounts, the following alternative-form hypotheses were developed for experiment two:

H3: The proportion of auditors having an above-criterion (as defined in experiment one) portion of judgment (misstatement risk) variation attributable to an interaction of gross sales material change and net accounts receivable material change will be significantly greater than chance.

H4: For auditors having an above-criterion portion of judgment (misstatement risk) variation attributable to the expected interaction, such interaction will be disordinal with the following specific attributes:

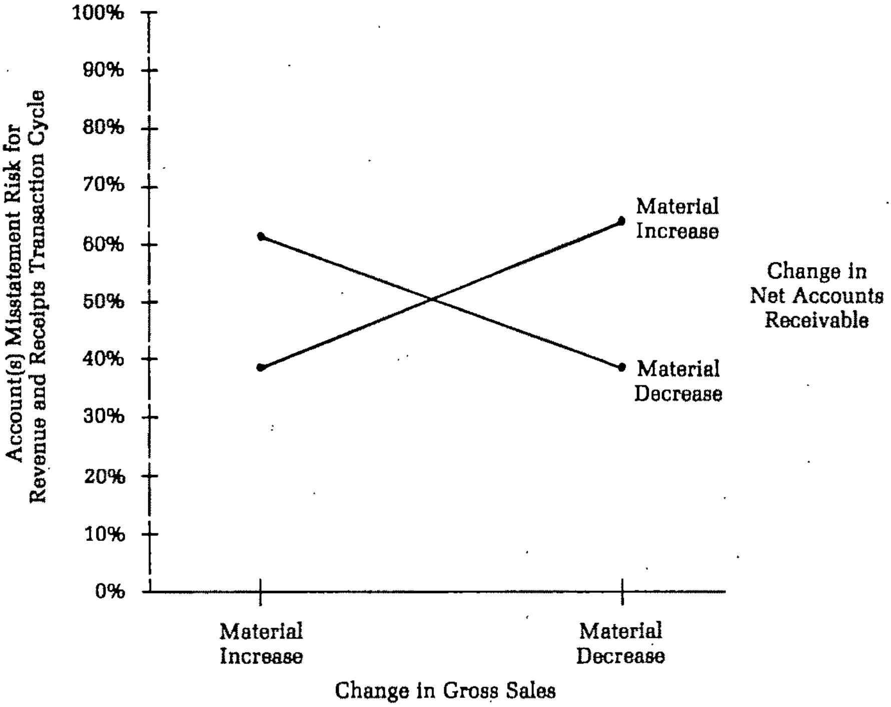
a. Misstatement risk assessments will be significantly larger when net accounts receivable has increased materially than when it has decreased materially, given that gross sales (in both instances) has decreased materially.
b. Misstatement risk assessments will be significantly smaller when net accounts receivable has increased materially than when it has decreased materially, given that gross sales (in both instances) has increased materially.

Subjects

Twenty-two audit seniors with three to four years of audit experience participated in experiment two, none of which partcipated in experiment one. These subjects were employed by the same national CPA firm and participated at one of the firm's offices.

Figure 2

Example of a Disordinal Interaction in the Context of Experiment Two



Research Design

The research design was a one-half fractional replication of the factorial manipulation of five cues, each at two levels, involving trial balance changes for selected accounts. Similar to experiment one, an ANOVA model was computed for each auditor-subject's risk assessments that estimated all five main effects and ten two-way interactions. Again, since such models are determined fully, variation in explained judgment was computed using the same method and the same arbitrary criterion for determining the significance of a term's explained judgment variation was employed in experiment two as in experiment one (i.e., ≥4 percent of total judgment variation).

Materials

A hypothetical audit case was developed in which, as discussed above, configural information processing would be consistent with auditing knowledge specific to evaluation of audit planning analytical procedures. The specific context was assessment of misstatement risk for revenue-and-receipts cycle accounts given changes in selected trial balance figures between the current period (unaudited) and the prior period (audited). The case included background information and information concerning both possible changes in the auditee's economic circumstances and its business and ac-

Exhibit 2

Example of an Analytical Procedure Comparing Selected Trial Balance Figures from Experiment Two

Account	*Change*
A. Sales (gross)	+25.6%
B. Cash Balances	+19.8%
C. Short-Term Investments	−20.2%
D. Accounts Receivable (net)	−24.3%
E. Inventories	+23.6%
F. Fixed Assets	+0.7%

counting policies (presented in appendix B). Specimen stimuli presentations also were included.

The subjects were told that they would be given a series of comparisons, prepared by an auditor on their staff, of the selected trial balance figures and would be asked to assess the following misstatement risk (elicited on a 100-point scale where 0 was no risk and 100 was maximum risk):

> Given the information obtained from client personnel, together with the comparison of selected trial balance figures (above), assess the risk that account(s) within Agco's REVENUES AND RECEIPTS CYCLE could be materially misstated.

An example of a trial balance figure comparison is presented in exhibit 2. Gross sales and net accounts receivable were intended to be the related revenues and receipts cycle accounts and are the constituent variables of the interaction expected in hypothesis H3. The other four accounts were intended to be from transaction cycles other than revenues and receipts (and thus, not directly related to accounts within the revenues and receipts cycle). Five of the six accounts (A, B, C, D, and E in exhibit 2) were factorially manipulated at two levels each ("material increase" or "material decrease"), and one account (F in exhibit 2) was held constant ("no material change").[10]

Procedures

The procedures for experiment two were similar to those described earlier for experiment one.[11] On average, subjects took 18.5 and 13.5 minutes per training and testing section, respectively.

Results

H3: Expected Interaction. Twenty-one of the 22 auditor-models (95.5 percent) attributed above-criterion judgment variation to the interaction expected in hypothesis

---

[10] Rather than verbal labels, material increase (decrease) was reported to subjects as a normally distributed random variable (percent) with mean 25 (−25) and standard deviation 6. No material change was reported as a normally distributed random variable (percent) with mean 0 and standard deviation 0.15.

[11] The primary procedural differences were that experiment two's training section involved four manipulations of the experiment section case, and the only decision aid available to subjects was an electronic file of trial balance comparisons already evaluated.

Table 3
Summary of Auditor Judgment Models: Experiment Two

	Judgment Models Incorporating:				
	Only Above-Criterion Terms			*All Terms*[1]	
Source[2]	*Mean*[3]	*Std. Dev.*	n[4]	*Mean*[3]	*Std. Dev.*
A	8.09	4.44	8	3.72	4.34
B	11.85	5.80	3	2.16	4.38
C	8.24	3.37	5	2.88	3.51
D	7.00	4.80	8	3.31	4.08
E	9.00	3.54	10	4.61	4.81
All Main Effects	14.37	9.64	20	16.67	8.82
AB	10.15	5.68	13	6.60	6.21
AC	11.89	6.70	14	7.79	7.67
AD	39.85	25.92	21	38.08	26.63
AE	13.80	10.54	10	7.15	9.10
BC	10.95	6.23	11	6.45	6.38
BD	8.99	2.97	4	2.30	3.53
BE	9.21	5.59	9	4.41	5.40
CD	9.89	10.70	6	3.72	7.20
CE	7.39	4.12	4	2.58	3.03
DE	11.42	5.99	7	4.25	5.99
All Interactions	76.88	8.51	22	83.33	8.82
Residual	10.05	4.12	22		

[1] The number of auditor-models (n) equals 22.
[2] Sources are keyed to procedures in exhibit 2. The expected interaction is AD.
[3] Judgment variation attributed to the respective source.
[4] The number of auditor-models for which the specified term is above-criterion.

H3 (AD in table 3). This frequency is significantly greater than chance (a 95 percent confidence interval for this frequency is 85.5 percent to 100 percent) and thus is consistent with hypothesis H3. The mean judgment variation attributed by the 21 auditor-models to the expected interaction was 39.9 percent (a 95 percent confidence interval for this mean is 28.09 percent to 51.61 percent).[12]

---

[12] An analysis of these results' sensitivity to the arbitrary criterion (threshold) used in determining significance of judgment variation attributed to the expected interaction indicated the following:

	Judgment Variation Criterion				
Expected Interaction	≥0%	≥4%	≥10%	≥20%	≥30%
Percent of subjects	100.0	95.5	86.4	68.2	54.5
Mean judgment variation	38.0	39.9	43.2	51.4	58.2
Total number	22	21	19	15	12
Number in expected form	21	21	19	15	12

Using the same testing methods described in the paper, both hypotheses H3 and H4 are supported using any criterion within the range of this sensitivity analysis (0 to 30 percent).

Table 4
Misstatement Risk Assessments: Experiment Two

	Levels of the Expected Interaction			
	II	*ID*	*DI*	*DD*
Mean (n=20)	38.34	61.75	63.36	38.83
Standard Deviation	14.68	15.14	12.82	13.13

Note: The expected interaction is change in gross sales (A) by change in net accounts receivable (D). The levels are: II=both A and D have materially increased; ID=A has materially increased and D has materially decreased; DI=A has materially decreased and D has materially increased; and DD=both A and D have materially decreased.

H4: Form and Nature of the Expected Interaction. Although table 3 indicates that 21 of the 22 auditor-models attributed above-criterion judgment variation to the expected interaction (AD in table 3), visual inspection revealed that one such interaction did not have the expected specific attributes.[13] The form and nature of the interaction is as expected for the remaining 20 auditor-models (90.9 percent), which supports the initial part of hypothesis H4 (also see fn. 12). The mean risk assessments for each level of this interaction are presented in table 4. A one-way repeated measures ANOVA indicated that these cell means are significantly different ($F(3,57)=55.7$; $p<0.01$). Tukey's test over the levels of interaction indicates that assessments of misstatement risk are significantly larger when net accounts receivable has increased than when it has decreased given that gross sales (in both instances) has decreased (DI and DD in table 4). In addition, assessments of misstatement risk are significantly smaller when net accounts receivable has increased than when it has decreased given that gross sales (in both instances) has increased (II and ID in table 4). Together, these results verify the two attributes specified in the later parts of hypothesis H4 and thus confirm that the predicted disordinal interaction was of the expected form (also see fn. 12).

Case Design Validity Check. A validity check analogous to that performed for experiment one also was performed for experiment two. As previously indicated, 20 of the 22 auditor-models (90.9 percent) attributed above-criterion judgment variation to the expected disordinal-form interaction. Of the nonhypothesized above-criterion interactions, AC (see table 3) had the greatest frequency (14 of the 22 auditor-models). This frequency is significantly smaller than that for the expected interaction ($Z=2.16$; $p<0.03$). Further, the mean judgment variation attributed by the 20 auditor-models to the expected interaction (41.5 percent), is significantly greater than that attributed to this nonhypothesizd interaction (11.9 percent; $t(32)=4.16$; $p<0.01$).

IV. Discussion and Concluding Remarks

In this paper we report the results of two experiments in which auditors' configural processing was investigated. The evidence indicates that a high proportion of the auditor-subjects configurally processed the available information (55.1 percent in experiment one and 90.9 percent in experiment two). The mean judgment variation attributable to the *single* hypothesized interaction in experiment one is three times that of any single interaction in prior studies. Further, the judgment variation attributable to interactions in experiment two, in contrast to prior studies, was substantially larger than that attributable to main effects. Moreover, the results confirm that the form and nature of the auditor-subjects' configural information processing were predictable based on auditing knowledge specific to the task being performed. These results highlight the importance of using context- and task-specific knowledge to identify *ex ante* those information processing strategies that judges are likely to employ. Such identification, in turn, facilitates the design of experimental materials and procedures as well as interpretation of the experimental results. As observed by Brown and Solomon (1990), without such prior identification, interpretation of observed effects is problematic.[14]

---

[13] Although the expected interaction for this auditor-model was disordinal, its specific attributes were opposite of those expected (suggesting that this subject mistakenly responded using a "reliability" rather than "risk" scale).

Together with Brown and Solomon (1990), this study provides a different view of auditor configural information processing than previously has been reported.[15] Over a wide variety of tasks (e.g., assessment of control risk, evaluation of audit procedure evidence, analysis of financial statement account comparisons) for which task- and context-specific auditing knowledge implies that it is appropriate, many auditors apparently are able to configurally process information. As with all laboratory experiments, however, limitations to generality exist. In the present experiment, subjects were obtained from a single audit firm and materials and procedures were employed that differ from earlier studies. Microcomputer administration (instead of paper and pencil), decision aids (intended to reduce memory load and control extraneous judgment variation), and training procedures (to mitigate subject reactivity) all differ from previous studies and could have contributed to the present finding of greater configural information processing than in prior studies.

The present study and Brown and Solomon (1990) also reveal that the *magnitude* of configural information processing may be task- and context-specific. The mean judgment variation attributable to the expected interactions was 7.4, 12.2, and 41.4 percent in Brown and Solomon (1990) and the present experiments one and two, respectively. Comparing the two ordinal tasks, anecdotal evidence suggests that the mediating effect on risk assessment of the rate of substitution between compensating controls is perceived by auditors to be smaller than that of the rate between alternative audit procedures. In addition, considering the disordinal task, the "reversal" effect on risk assessments of opposite changes in related financial statement accounts that generally change in the same direction is greater than the mediating effects due to rates of substitution.

---

[14] Although many of the nonhypothesized interactions observed in experiments one and two can be "interpreted" *post hoc*, the problematic nature of such interpretations is the major reason for not doing so.

[15] Mayper et al. (1989) subsequently have reported a significant disordinal interaction in ANOVA models of internal control weakness materiality judgments for 18 of their 34 (52.9 percent) auditor-subjects. Their interaction involved type of asset and type of control weakness. Specifically, reflecting the notion that general control objectives are differentially important to specific control subsystems, lack of authorization control weaknesses were judged relatively more material if the weakness affected noncash assets rather than cash, but separation-of-duties control weaknesses were judged relatively more material if the weakness affected cash rather than noncash assets. However, the specific auditing knowledge that would manifest such a disordinal interaction was not specified a priori by the researchers and thus the interaction was not anticipated.

Finally, there are differences in the *proportions* of auditors exhibiting the expected configural information processing strategies between the three experimental contexts. These proportions varied from 40.5 percent in Brown and Solomon (1990) to 55.1 percent in the present experiment one to 90.1 percent in the present experiment two. We speculate that such variability may be due, in part, to requisite knowledge differences across the three experimental tasks. For example, knowledge of the accounting model and the interrelationships among primary financial statement accounts (required for the present experiment two) is taught in college classes and is fundamental to auditing. Knowledge concerning substitutable audit procedures or controls, however, is more advanced. Consequently, the proportions of audit seniors having the former knowledge (and thus exhibiting configural information processing) might be higher than the latter knowledge. These speculations concerning differences in magnitude and frequency of configural information processing profitably could be investigated in future research.

Appendix A
Accounts Receivable Audit Procedures Case From Experiment One

Assume you are a senior-level auditor and that one of your clients is Agco, Inc. Agco, a large processor of corn and soybean products, is a privately held company with publicly traded bonds that require audited financial statements prepared in accordance with GAAP. The company has not presented significant auditing problems during your firm's five-year tenure as its public auditor, and has maintained moderately good internal accounting control systems. During the past two years, Agco has been computerizing its accounting and information systems.

Currently, you are performing Agco's 1988 year-end audit and are evaluating evidence provided by the audit of accounts receivable. During engagement planning, the inherent risk that a material error could arise in accounts receivable was assessed to be moderate for both existence and valuation. For [16] randomly ordered cases, you will be presented with the results from a portion of the accounts receivable audit program performed by your staff auditors. For each case, you will be asked to assess the risk that accounts receivable, given the audit evidence-to-date, could be materially misstated AS A RESULT OF recorded accounts NOT EXISTING at year-end.

Additional accounts receivable audit procedures information:

A. The sample for accounts receivable confirmation was drawn to obtain 95 percent reliability estimates. WHEN indicated as being COMPLETED, responses have been received (through second and third requests) from 100 percent of the sample.
B. Verification of accounts receivable sales transactions and observation of subsequent collections is a SEPARATE audit procedure from the confirmation of those accounts. The sample for this procedure (also drawn to obtain 95 percent reliability estimates) is INDEPENDENT of that for the confirmations.
C. The analytical procedures used to verify the adequacy of the allowance for bad debts are to compare the current year's number of days accounts outstanding with that of previous years, and to compare the allowance as a percentage of accounts receivable with that of previous years.
D. The aged accounts receivable listing is reviewed for material receivables from affiliates, directors, and other related parties; notes and long-term receivables; and accounts with significant credit balances.

Appendix B
Financial Statement Analytical Comparisons Case From Experiment Two

Assume you are a senior-level auditor and that one of your clients is Agco, Inc., a privately held company that has debt covenants requiring audited financial statements prepared in accordance with GAAP. Previously, the company was audited by another CPA firm (whose audit opinions have been unqualified for the past five years), but Agco had its audit services competitively re-bid. This is your firm's first year as Agco's auditor. A large processor of agricultural products, Agco has processing facilities in central Illinois and eastern Iowa, and a corporate headquarters in Illinois. Agco sells its products primarily to a variety of food processors located world-wide. Agco uses a standard process costing system which periodically reports to management actual results compared with a flexible budget, as well as a small, but competent staff of internal auditors.

Through discussions with the prior auditor and client personnel, you have obtained the following information related to the previous five years:

1. There have been no changes in accounting policies or methods.
2. There were no contentious accounting issues with the prior auditor.
3. Although financial information generated by its accounting system has been timely and adequate, there have been a few material recording errors (e.g., transpositions, misclassifications, and cut-offs) requiring audit adjustments. These errors have varied from year-to-year in terms of the specific transaction cycle(s) affected. Only ONE such error has been found in revenue and receipts cycle accounts.
4. Management has good attitudes toward financial reporting.
5. Management is highly competent and honest, and its turnover is very low.
6. There are no material amounts of consignment or installment sales.
7. There are no material amounts of intercompany or related-party sales.
8. The company's credit policies are unchanged from prior years.
9. There are no material amounts of trade notes receivable.
10. The pattern of receivable write-offs have not changed from prior years, and there has been no change in recoveries of written-off receivables.
11. Aging of receivable balances has not changed from prior years.
12. Agco has had strong boundary controls over the capture of exchange data within the revenue and receipts cycle, and these controls are unchanged from prior years.

Currently, you are performing preliminary engagement planning for Agco's 1989 year-end audit. As a result of the date at which your firm assumed its responsibilities, you have been unable to perform interim audit work. The audit manager initially has assessed the engagement risk as low; inherent risk assessments are yet to be made. For [16] randomly ordered cases, you will be presented with comparisons of selected trial balance figures for the current year and those from the prior year. For each case you will be asked to assess the risk that account(s) within the REVENUES AND RECEIPTS CYCLE could be materially misstated.

References

AICPA. 1983. *Statement on Auditing Standards No. 47.* New York: AICPA.

Ashton, R. H. 1974. An experimental study of internal control judgments. *Journal of Accounting Research* 12 (Spring): 143–57.

———. 1982. *Human Information Processing in Accounting.* Sarasota, FL: AAA.

———. 1983. *Research in Audit Decision Making: Rationale, Evidence, and Implications.* Vancouver, British Columbia: Canadian Certified General Accountant's Research Foundation.

———, and P. R. Brown. 1980. Descriptive modeling of auditors' internal control judgments: Replication and extension. *Journal of Accounting Research* 18 (Spring): 269–77.

———, and S. S. Kramer. 1980. Students as surrogates in behavioral accounting research: Some evidence. *Journal of Accounting Research* 18 (Spring): 1–15.

Brown, C., and I. Solomon. 1990. Auditor configural information processing in control risk assessment. *Auditing: A Journal of Practice & Theory* 10 (Fall): Forthcoming.

Camerer, C. 1987. MCPL learning of disordinal interactions, and the validity of expert judgment. Working paper, University of Pennsylvania.

Castellan, N. J. 1977. Decision making with multiple probabilistic cues. In *Cognitive Theory* (vol. 2) edited by N. J. Castellan, D. B. Pisoni, and G. P. Potts. Hillsdale, NJ: Erlbaum.

Davis, J., and I. Solomon. 1989. Experience, expertise and expert-performance research in public accounting. *Journal of Accounting Literature* 8: 150–64.

Dawes, R. 1988. *Rational Choice in an Uncertain World*. New York: Harcourt Brace Jovanovich.

Einhorn, H. J., and R. Hogarth. 1981. Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology* 32: 52–88.

———, D. N. Kleinmuntz, and B. Kleinmuntz. 1979. Linear regression and process tracing methods of judgments. *Psychological Review* 86 (September): 465–85.

Frederick, D. M., and R. Libby. 1986. Expertise and auditors' judgments of conjunctive events. *Journal of Accounting Research* 24 (Autumn): 270–90.

Goldberg, L. R. 1968. Simple models or simple processes? Some research on clinical judgments. *American Psychologist* 23 (July): 483–96.

Hall, C., P. Yetton, and I. Zimmer. 1982. The assessment of payroll internal control systems and auditors' experience, tolerance of ambiguity, and dogmatism. *Australian Journal of Management* (June): 49–60.

Hamilton, R. E., and W. F. Wright. 1982. Internal control judgments and effects of experience: Replications and extensions. *Journal of Accounting Research* 20 (Autumn, Part II): 756–65.

Hammond, K. R., and D. A. Summers. 1965. Cognitive dependence on linear and nonlinear cues. *Psychological Review* 72 (May): 215–34.

Hoffman, P. J. 1968. Cue-consistency and configurality in human judgment. In *Formal Representation of Human Judgment* edited by B. Kleinmuntz. New York: Wiley.

Leon, M., G. C. Oden, and N. H. Anderson. 1973. Functional measurement of social values. *Journal of Personality and Social Psychology* 27 (September): 301–10.

Libby, R. 1981. *Accounting and Human Information Processing*. Englewood Cliffs, NJ: Prentice-Hall.

Mayper, A., M. Doucet, and C. Warren. 1989. Auditors' materiality judgments of internal accounting control weaknesses. *Auditing: A Journal of Practice & Theory* 9 (Fall): 72–86.

Mock, T., and A. Wright. 1982. Evaluating the effectiveness of audit procedures. *Auditing: A Journal of Practice & Theory* 2 (Fall): 33–43.

Newell, A., and H. A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Payne, J. 1982. Contingent decision behavior. *Psychological Bulletin* 92 (September): 382–402.

Rorer, L., P. J. Hoffman, H. R. Dickman, and P. Slovic. 1967. Configural judgments revealed. *Proceedings of the 75th Annual American Psychological Association Convention*: 195–96.

Slovic, P. 1969. Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology* 53 (August): 255–63.

———, B. Fischhoff, and S. Lichtenstein. 1977. Behavioral decision theory. *Annual Review of Psychology* 28: 1–39.

Trotman, K. T., P. W. Yetton, and I. R. Zimmer. 1983. Individual and group judgments of internal control systems. *Journal of Accounting Research* 21 (Spring): 286–92.

Wright, A., and T. Mock. 1985. Toward a contingency view of audit evidence. *Auditing: A Journal of Practice & Theory* 5 (Fall): 91–100.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 120–139

Information Acquisition and Resource Allocation Decisions

Ramji Balakrishnan
University of Iowa

SYNOPSIS: Firms' investment practices often seem to be suboptimal; for example, they may ration capital or underinvest in risky projects. Recent research (e.g., Antle and Eppen 1985, Antle and Fellingham 1990, Harris et al. 1982) argues that these practices might be optimal responses to contracting frictions that arise from information asymmetry among the members of the firm.[1] Analyzing a two-person firm, these studies endow the manager with pre-decision information about the returns from a given project and examine the owner's optimal response to the information asymmetry. This article expands the problem in two ways. First, it considers the choice between two competing projects. Second, it explicates the source of the manager's information superiority: proximity to the firm's operations advantage the manager in both acquiring and knowing the cost-benefit of additional information pertinent to project choice. Because of the manager's advantage, the owner cannot infer why the project was selected.[2] The owner's response to the information asymmetry must optimize over both the information-acquisition and resource-allocation decisions.

This article considers a model in which the manager must choose between a risk-free and a risky project. The owner delegates project selection

---

[1] Analyzing a one-period model in which the risk-neutral manager has perfect pre-contract information, Antle and Eppen (1985) show that capital rationing and organizational slack are optimal responses to the information asymmetry. Extending Antle and Eppen to a two-period setting, Antle and Fellingham (1990) show value to long-term contracting (maintaining a history of budget reports). Examining a model of resource acquisition and its allocation among the many divisions, Balakrishnan (1990) shows that resource rationing at both the acquisition and the allocation stages is optimal.

[2] Although this article is couched in terms of resource allocations, other examples of contracting frictions that arise because the agent is better equipped to collect and interpret information abound. See Demski and Sappington (1987) for a general analysis of the problem of "delegated expertise."

I thank the participants at the University of Iowa accounting workshop for helpful comments. Special thanks to Douglas DeJong, John Fellingham, Bruce Johnson, and Rick Young, and particularly to two anonymous referees for their detailed comments. Barbara Yerkes provided editorial assistance.

Submitted October 1989.
Revisions received March and August 1990.
Accepted August 1990.

because of the manager's pre-contract information superiority and because only the manager can obtain post-contract additional, costly information. Delegating project choice, however, creates a coordination problem because of the owner's inability to observe whether the manager expended effort on information search and because the value of additional information depends on the manager's pre-contract information (i.e., on the manager's *type*). Using the observed variables (project choice and realized profit), the optimal contract must (1) motivate a subset of manager types to select the project after expending effort on information search, and (2) motivate the complementary set to condition project selection on their pre-contract information alone.

The model considered here is related to that of Demski and Sappington (1987) which considers issues involved in contracting with an "expert" who is uniquely qualified to acquire pertinent information, and Lambert (1986) applying this concept to a project-selection setting.[3] Pre-contract asymmetry about the value of search is the crucial difference between previous research and this study.

Absent information asymmetry (in the *first-best* solution), the information-acquisition problem is separable from the resource-allocation problem. The owner asks for search only from those manager types for whom the value to search exceeds cost. (The set of such manager types is convex because the value to search is concave in manager type.) Also, the manager uses all available information to choose the more profitable project.

However, the owner cannot attain this first-best solution without information about manager type and information search. The two problems cannot be decomposed. As a result, the optimal (second-best) contract always induces distortions in information acquisition and, for the case of the risk-averse manager, induces overinvestment in the risky project. Furthermore, communication of the search result is strictly valuable.

The results suggest that optimal responses to information asymmetry among the members of the firm could result in practices that seem to disregard the value of information search and to use acquired information suboptimally. The results also lend support to the use of accounting reports, such as budgets, in managerial performance evaluation.

Key Words: *Agency models, Information search, Project selection, Communication.*

THIS article models a two-person firm wherein the manager is advantaged both in acquiring and knowing the cost-benefit of additional information pertinent to the choice between a risky and a risk-free project. The owner's inability to observe whether the manager expended effort on information search prevents separation of the information-acquisition problem from the resource-allocation problem. Using the observed variables (project choice and realized profit), the optimal contract must motivate a subset of manager types to select the project after expending effort on information search and motivate the complementary set to condition project selection on their pre-contract information alone. Three main results obtain.

---

[3] Conroy and Hughes (1987) consider a model in which the owner collects the additional information. Information acquisition has also been modeled in the context of adoption of innovations. For example, in a static decision theoretic setting, Jensen (1982) examines incentives for the firm to collect information about expected profitability.

First, if the manager is risk-neutral, the set of manager types instructed to acquire information shifts to the right, relative to the same set in the first-best contract. The shift occurs because the optimal contract exploits the interaction between the manager's pre- and post-contract information. Increases in managerial risk aversion cause a strict reduction in the set of manager types asked to collect additional information. In the limit, the set is empty.

Second, if the manager is risk-averse, the optimal contract also induces distortions in the investment decision by calling for overinvestment in the risky project. The intuition is that the risk-free project is the *ex ante* choice of the "low" manager types. To these manager types, search incentives are provided through rewards for selecting the risky project. Hence, instructing overinvestment in the risky project enhances these manager types' incentives to expend effort on information search. As a result, the owner's expected payout decreases and offsets the higher expected project cost from the overinvestment.

Third, communication of the search result (e.g., bottom-up budgeting) is strictly valuable. If the manager is risk-neutral and there were no communication, the optimal contract induces underinvestment in the risky project. However, communication allows for the same incentives to be provided without underinvestment. Hence, communication has strict value. A similar rationale operates for a risk-averse manager. Specifically, the risky project is the *ex ante* choice of the "high" manager types. For such manager types, the optimal contract provides search incentives through search results that induce the manager to select the risk-free project. Hence, absent communication, the optimal contract induces underinvestment in the risky project. Again, communication is valuable because it allows the owner to provide search incentives as if the contract called for underinvestment.

I. The Model

The model consists of the owner of the firm and one manager.[4] The risk-neutral owner's utility is defined over monetary receipts only. The effort-averse manager's utility is defined over monetary receipts and effort. The manager's utility is additively separable in its components.

The owner employs the manager to choose between a risk-free and a risky project. Both projects have the same output value, X. The risk-free project costs C_m to implement. The risky project is either a success with cost C_s or a failure with cost C_f $(C_f > C_m = X > C_s)$. The following paragraphs detail the sequence of events that further define the owner's problem.[5] It is helpful to refer to figure 1, which summarizes the sequence of events.

---

[4] Only one manager of unknown type is assumed to exclude the optimality of some kind of a competitive scheme as in Demski and Sappington (1984).

[5] The regulation literature explores optimal policies for controlling the firm's ability to extract consumer surplus because of its private information about costs (e.g., Baron and Myerson 1982; Laffont and Tirole 1986).

Pre-contract asymmetry concerns the manager's skill (i.e., the manager's *type*). Index manager type by $i \epsilon [i_{min}, i_{max}]$ and denote the owner's prior beliefs about manager type by $\phi(i)$. Manager type is relevant because the *ex ante* probability of success in the risky project (denoted $P_i(C_s)$) is strictly increasing in manager type; $dP_i(C_s)/di > 0$. Without loss of generality, confine the analysis to truth-revealing contracts (Myerson 1982).

At contracting time, the owner has the option of contracting with the manager to conduct an unobservable information search and then choose a project. If information search is instructed, the manager incurs disutility (denoted a, with $a > 0$) to conduct the search. The search reveals additional information pertinent to the profitability of the risky project. The manager's pre- and post-contract information interact because the manager's skill (type) and the search result together determine the profitability of the risky project. Formally, index the random state of nature by j and let $P(j)$ denote the common ex ante beliefs about j. Then, $P_i(C_s) = \int_j P_{ij}(C_s) P(j) dj$, where $P_{ij}(C_s)$ denotes the updated probability of success in the risky project. Interaction between manager type and the state of nature favorably affects the probability of success in the risky project; $dP_{ij}(C_s)/dj > 0$ and $dP_{ij}(C_s)/di > 0$. That is, the more skilled the manager, the more the states in which the risky project has lower expected cost. Further, for a given state of nature, the more highly skilled the manager, the lower the expected cost of the risky project. For compactness, $P_{ij}(C_s)$ will be written as P_{ij} where there is no confusion.

If information search is instructed, the manager communicates the private search result to the owner before selecting a project. Hence, the owner's instructions on project selection can depend both on manager type and the reported search result. Denote the instructed investment cutoff by r_{ii}. If $P_{ij} \leq r_{ii}$, the manager is instructed to choose the risk-free project, and to choose the risky project otherwise.

Of course, the owner could also instruct the manager to condition project selection on the manager's pre-contract information only. Instructing information search is only an option. With either option (project selected with or without search), the realized cost of the chosen project is publicly observed. Then, the manager is paid as promised.

The Owner's Problem

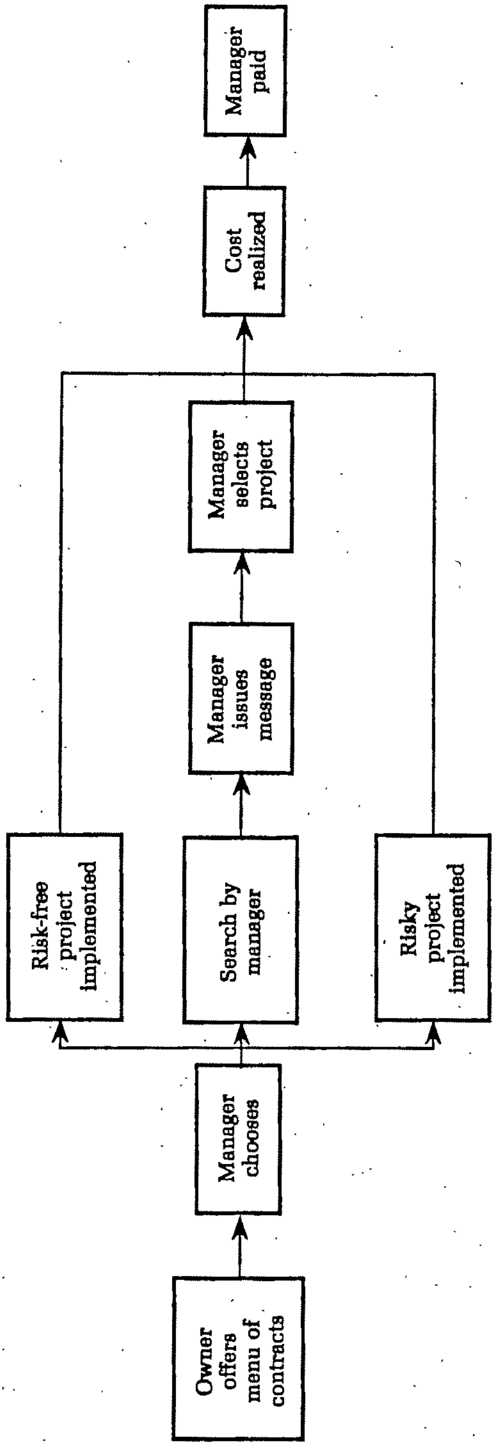
Formulating the owner's problem is considerably simplified because:

Observation 1: In the optimal contract, the set of manager types instructed to acquire information is convex.

Proof: Assume to the contrary that the optimal contract does not ask the type i manager to search but instructs manager types $i \pm \epsilon$ to search. If the type i manager is instructed to choose the risk-free project without search, truth telling would be destroyed. Because some manager types $i' < i$ acquire information, the expected utility to such manager types is strictly increasing in manager type. Manager types $i' < i - \epsilon$ would therefore claim to be type i; they are certain to obtain higher utility. Contradiction.

Constraints similar to those here are used to make it in the best interest of the firm to disclose its costs truthfully. These models, however, use an objective function that is a weighted average of consumer surplus and firm profit. Because I wish to explore relationships within the firm, the owner is granted property rights to the output from the project.

Figure 1
Time Line of Events



> Next, assume that the type i manager is instructed to choose the risky project without search. Note that instructing search is profitable (from the owner's perspective) for manager types $i \pm \epsilon$. Then, with all else constant, offer the type i manager the contract offered to the type $i-\epsilon$ manager. This action increases the owner's expected utility. Contradiction. The set of manager types instructed to acquire additional information must be convex.■

Hence, the owner groups manager types into three classes at most. The class 1 manager types are instructed to choose the risk-free project without search. The class 2 manager types are instructed to select the project after expending effort to acquire additional information. The class 3 manager types are instructed always to choose the risky project. Of course, each class could consist of many manager types or be a null set. Formally, denote the cutoffs used to define manager classes by M_l and M_h. Then, given the information available to the owner, the payment scheme is denoted by $I1_i$, $I2_{ijk}$, and $I3_{ik}$, respectively for the three manager classes; subscript k indexes the realized cost of the chosen project.

Finally, for convenience in writing the program, let $EU(i)$ denote the expected utility of a type i manager who is honest and obedient. Note that the expression for $EU(i)$ would vary in manager class as the instructions differ by manager class (see the set of (M) constraints in the program). Next, initialize the manager's reservation utility to zero and set $U(0)=0$. Then, the owner's problem is to choose $I1_i$, $I2_{ijk}$, $I3_{ik}$, M_l, M_h, and r_{is} to maximize:

$$
\begin{aligned}
\text{(P)}\quad & \textstyle\int_i [X - C_m - I1_i]\phi(i<M_l)\,di \\
& + \textstyle\int_i [\int_j \{w\Sigma_k (X - C_k - I2_{ijk})P_{ij}(C_k) \\
& \quad + (1-w)(X - C_m - I2_{ijm})\}P(j)\,dj]\phi(M_l \le i \le M_h)\,di \\
& + \textstyle\int_i [\Sigma_k (X - C_k - I3_{ik})P_i(C_k)]\phi(i>M_h)\,di
\end{aligned}
$$

subject to:

$$U(I1_i) \ge 0 \quad \forall i<M_l \tag{M1}$$

$$\textstyle\int_j [w\Sigma_k U(I2_{ijk})P_{ij}(k) + (1-w)U(I2_{ijm})]P(j)\,dj - a \ge 0 \quad \forall M_l \le i \le M_h \tag{M2}$$

$$\Sigma_k U(I3_{ik})P_i(k) \ge 0 \quad \forall i>M_h \tag{M3}$$

$$EU(i) \ge U(I1_{i'}) \quad \forall i \text{ and } \forall i'<M_l; \tag{TT i1}$$

$$EU(i) \ge \Sigma_k U(I2_{i'j'k})P_i(k) \quad \forall i,\ M_l \le i' \le M_h,\ j': P_{i'j'} > r_{i's} \tag{TT i2.1}$$

$$EU(i) \ge U(I2_{i'j'm}) \quad \forall i,\ M_l \le i' \le M_h,\ j': P_{i'j'} \le r_{i's} \tag{TT i2.2}$$

$$
\begin{aligned}
& EU(i) \ge \textstyle\int_j \text{Max}_{j'} \{\Sigma_k U(I2_{i'j'k})P_{ij}(k),\ U(I2_{i'j'm})\}P(j)\,dj - a \\
& \forall i,\ M_l \le i' \le M_h
\end{aligned}
\tag{TT i2.3}
$$

$$EU(i) \ge \Sigma_k U(I3_{i'k})P_i(k) \quad \forall i \text{ and } \forall i'>M_h, \tag{TT i3}$$

with $w=1$ for $P_{ij}>r_{is}$ and $w=0$ for $P_{ij} \le r_{is}$. The RHS of constraint (TT i2.3) reflects that the manager's message and conditional project choice would maximize personal expected utility.

The set of (M) constraints is the market constraint for the three classes. For each i, the set of (TT) constraints ensure that truth telling yields expected utility (i.e., $EU(i)$)

that weakly exceeds the expected utility from all possible lying strategies. Hence, the manager will reveal the truth both for type and for the search result (if any).

Some of the (TT i2) constraints deserve special mention. Search must be induced for all the class 2 manager types. That is, for $M_l \leq i \leq M_h$,

$$\begin{aligned}&\textstyle\int_j [w\Sigma_k U(I2_{ijk})P_{ij}(k)+(1-w)U(I2_{ijm})]P(j)dj-a \\ &\quad\geq \Sigma_k U(I2_{ij'k})P_i(k) \quad \forall j':P_{ij'}>r_u;\end{aligned} \tag{IS 1}$$

$$\begin{aligned}&\textstyle\int_j [w\Sigma_k U(I2_{ijk})P_{ij}(k)+(1-w)U(I2_{ijm})]P(j)dj-a \\ &\quad\geq U(I2_{ij'm}) \quad \forall j':P_{ij'}\leq r_u.\end{aligned} \tag{IS 2}$$

Next, given truthful reporting of type and given signal j from the search, the manager must be induced to truthfully report the search result. That is, for j such that $P_{ij}>r_u$,

$$\Sigma_k U(I2_{ijk})P_{ij}(k)\geq U(I2_{ij'm}) \quad \forall j':P_{ij'}\leq r_u; \tag{SR 1}$$

$$\Sigma_k U(I2_{ijk})P_{ij}(k)\geq \Sigma_k U(I2_{ij'k})P_{ij}(k) \quad \forall j':P_{ij'}>r_u; \tag{SR 2}$$

and for j such that $P_{ij}\leq r_u$,

$$U(I2_{ijm})\geq U(I2_{ij'm}) \quad \forall j':P_{ij'}\leq r_u; \tag{SR 3}$$

$$U(I2_{ijm})\geq \Sigma_k U(I2_{ij'k})P_{ij}(k) \quad \forall j':P_{ij'}>r_u. \tag{SR 4}$$

The First-Best Solution

Before analysis of the above program, consider the solution with symmetric information (the *first-best* solution). Here, because the value of additional information is concave in manager type, the owner chooses cutoffs M_l^* and M_h^* to group manager types.[6] Manager types $i<M_l^*$ always choose the risk-free project, and manager types $i>M_h^*$ always implement the risky project (because $P_i(C_s)$ is increasing in manager type). Manager types $M_l^*\leq i\leq M_h^*$ select after expending effort to search. In particular, if r_f satisfies $C_s r_f+C_f(1-r_f)=C_m$, the class 2 manager chooses the risky project for all $P_{ij}>r_f$ and chooses the risk-free project otherwise. In return for expertise in implementing the project, the class 1 and class 3 manager types are paid their reservation wage of 0 (because $U(0)=0$). The class 2 manager type is paid a flat wage to offset the disutility connected with the search.

The above solution is not feasible in the presence of the assumed information asymmetry. First, search is costly and is unobservable. If the search entails zero disutility (i.e., if $a=0$), then paying the manager zero for all outcomes attains the first-best solution. Similarly, if search is observable (even if the search result is not), compensating managerial search at $U^{-1}(a)$ and paying zero for all realized costs removes all contracting frictions. Costly, unobservable search is key.

Second, there is pre-contract asymmetry about manager type (i.e., the value to search). If the risk-neutral manager's type is known, the owner sells the firm to the manager. Sequential rationality implies that the manager would not distort either the information-acquisition or the resource-allocation decision, relative to the first-best. If the manager with known type is risk-averse, the resultant model is similar to Lambert

---

[6] A formal proof of the claim that the value of information is concave in manager types can be obtained from the author.

(1986), who shows that risk must be imposed on the risk-averse manager to induce search. Also, via an example, Lambert shows that overinvestment in the risky project *could* occur. The distinction in the considered model is that manager type is unknown. Hence, the optimal contract is affected by the interaction between the manager's pre- and post-contract information. Even absent risk-sharing considerations, the owner's problem arises because the value to search depends on unknown manager type and hence requires the optimal contract to induce search only for a subset of manager types.

II. Analysis

This section begins by establishing the source for value to communication. Next, strict value to communication is established for the risk-neutral manager. Finally, the optimal contract is characterized for the risk-averse manager.

Proposition 1: The optimal contract never induces conditional underinvestment in the risky project. In the optimal contract, $r_{iu} \leq r_{if}$.[7]

Proof: See appendix.

The intuition is that the risk-free project provides a safe option to the manager. Communication permits the owner to provide the same compensation to the manager (through an outcome-independent transfer) if the risky project is implemented as would have been provided if the risk-free project had been implemented. Hence, communication allows the owner to provide search incentives as if the optimal contract called for underinvestment in the risky project. Thus, as in Lambert (1986), Proposition 1 shows weak value to communication. Strict value arises only if, absent communication, the optimal contract induces underinvestment in the risky project; $r_{iu} > r_{if}$ for at least one i. For the case of the risk-neutral manager, the following subsection establishes strict value to communication.

Risk-Neutral Manager

Begin by considering the case in which the owner does not observe project choice or realized cost. The resulting model is similar to that of Antle and Eppen (1985). The only source of friction between the risk-neutral players is pre-contract asymmetry on firm value, which is increasing in manager type.[8] The optimal contract is to sell the firm to the manager for price $V(M_i)$, where $V(M_i)$ is the value of the firm to manager type $M_i > M_i^*$.

However, if the owner observes realized cost (and thus project choice), selling the firm to the manager is not optimal. The intuition is that in the considered model, the manager does not have perfect pre-contract information. Perfect information on relative profitability requires managerial effort. Exploiting the interaction between the manager's pre- and post-contract information improves the owner's welfare, relative to selling the firm.

---

[7] The result readily extends to more general outcome distributions for the risky project or for more than one risky project. When a risk-free project is included in the menu, the manager never chooses it if one of the risky projects is preferred by the owner. However, with many risky projects, the risky project that is chosen could differ from the "best" risky project. Underinvestment of that form could persist even with communication.

[8] Proof of the claim that the value of the firm is increasing in manager type can be obtained from the author. Further, sequential rationality implies that manager types $M_i \leq i \leq M_i^*$ conduct the search and that conditional investment is first-best.

Proposition 2: Assume a risk-neutral manager. Then,

(1) In the optimal contract, the owner discriminates among manager types.
(2) In the optimal contract, the set of manager types instructed to acquire information shifts to the right, relative to the same set in the first-best solution.
(3) In the optimal contract, communication of the search result by the manager to the owner is strictly valuable.
(4) The optimal conditional investment decision is first best.

The intuition is that any contract selling the firm to the manager can be modified to increase the price by a nonzero amount and to offset the incremental price with a rebate for all "unfavorable" outcomes (select the risk-free project). Of course, the incremental price and the rebate could be chosen to yield the same expected utility to the lowest manager type buying the firm (i.e., to manager type M_l).[9] The modification is feasible because realized cost is public. Further, the modification preserves truth telling both for manager type and for the search result. However, the expected rebate is decreasing in manager type: interaction between manager type and the search result leads to the probability that an "unfavorable" outcome will obtain to decrease in manager type. The modification allows for implicit discrimination among manager types and, thus, increases the owner's expected utility, relative to selling the firm to the manager.

In fact, such a contract (offsetting incremental price through rebates for the risk-free project) is optimal. The manager's compensation is then (almost everywhere) the realized profit from the project less a charge that *varies* in the selected project.[10] Interpreting the charge as the imputed cost of invested capital, we see that the optimal contract resembles residual income (Horngren and Foster 1987, 873).

Discriminating among manager types, however, distorts information acquisition (relative to the first-best solution). Specifically, because the same contract is offered to all manager types, extracting information-rent accruing to the low manager types reduces the information-rent that could be extracted from the high manager types. Hence, the owner optimally trades off the foregone expected information rent from the low manager types with the greater expected information rents that could be extracted from the higher types; $M_l > M_l^*$. This trade off is generically termed the *upstream effect* by Mussa and Rosen (1978) and is the operative force in Antle and Eppen (1985) and Balakrishnan (1990).

Similarly, relative to the first-best solution, information acquisition at the high end of the manager types is distorted; $M_h > M_h^*$. The reason is that value to search arises from results that lead to the selection of the alternate to the default project (i.e., the project that would have been chosen with the pre-contract information alone). The default is the risky project for the highly skilled manager. Providing a rebate for the risk-free project increases the value of search to the high manager types. Hence, $M_h > M_h^*$ because of sequential rationality on the manager's part.

---

[9] Increasing the incremental price increases the range of unfavorable outcomes for which the rebate is provided. Then, because the agent's disutility for search is compensated by favorable outcomes only, the incremental price cannot be increased arbitrarily.

[10] Note that for $r_f < P_{ij} < r'$, the statement is not true. Hence, the modifier "almost everywhere."

Together, the above two features of the optimal contract imply that, relative to the first-best solution, the set of manager types instructed to search shifts to the right (i.e., $M_l > M_l^*$ and $M_h > M_h^*$). For the low manager types, the risk-free project is implemented more often than in the first-best solution. Also, in the second-best solution, manager types $M_h \geq i > M_h^*$ choose the risk-free project with positive probability. Thus, the optimal second-best contract results in the risk-free project being chosen more often, overall, than would have been chosen in the first-best solution.[11]

Discriminating among manager types leads to a strict value for reporting of the search result. The rebate for the risk-free project implies that the manager strictly prefers the risky project only for some $r' = P_U > r_f$. Absent communication, the optimal contract calls for underinvestment in the risky project. From Proposition 1, communication allows for the same search incentives to be provided without underinvesting in the risky project.[12] Hence, communication has strict value.[13]

Assuming a risk-averse manager, I next examine how the additional risk-sharing role for the contract affects the owner's optimal response to the information asymmetry and characterize the optimal response as distortions in the information-acquisition and resource-allocation decisions.

Risk-Averse Manager

Define $\lambda = -U''(\cdot)/U'(\cdot)$ as manager's Arrow-Pratt measure of risk aversion. Then, this subsection establishes that increases in the manager's risk aversion cause a strict reduction in the set of manager types asked to search. Further, strict value to communication and overinvestment in the risky project are established.

Proposition 3: Assume a risk-averse manager. Then:

(1) The set of manager types for whom the default is the risk-free project (class 2A types) are paid zero if the risk-free project is chosen. The optimal payment for the risky project is the solution to the system of equations defined by the (IS 2) constraint for manager type M_l and the constraint for truthful reporting of the search result at the cutoff probability for selecting the risky project (denoted r_{al}).

(2) For the manager types for whom the default is the risky project (class 2B types), the optimal contract is the solution to the system of equations defined by the (IS 1) constraint for manager type M_h, the truth-telling con-

---

[11] Note the distinction between *conditional* (after search) investment and *overall* investment (which includes cases wherein the project is selected without search).

[12] Note that because the cost outcomes are distinct for the two projects, the owner can perfectly verify the project choice *ex post*. Hence, there is strict value to communication even if project choice is observed by the owner. Communication enables the owner to better infer why the risky project was chosen. Specifically, the manager would communicate $r_f < P_U \leq r'$ and choose the risky project only if he or she has conducted the information search.

[13] Analyzing a general model of communication, Melumad and Reichelstein (1989) show that if the risk-neutral agent has perfect pre-contract information, communication has no value. However, they also show (1989, sec. 6; Melumad and Reichelstein 1987) that if the agent's decision results in a compression of private information and if the optimal contract requires different compensation for different messages, then delegated schemes without communication are inferior to schemes with communication. Such is the situation here. The set of possible search results is larger than the set of decisions. Also, the compensation for the risky project must differ from the compensation for the risk free project, for at least one search result. Hence, communication has strict value.

straint for the lowest manager type (denoted M_c) in this set of manager types and the constraint for truthful reporting of the search result at the cutoff probability for selecting the risky project (denoted r_{bs}).

(3) If the set of class 2A manager types is not empty, there is overinvestment in the risky project.

(4) If the set of class 2B manager types is not empty, there is strict value to communication of the search result.

(5) Increases in managerial risk aversion cause a strict reduction in the set of manager types instructed to acquire information.

The differences between the contracts in Propositions 2 and 3 highlight the different routes taken to motivate search. If the manager is risk-neutral, the owner's welfare is improved by lowering the effective cost of the *risk-free* project. Because the manager is the residual claimant, search is induced by sequential rationality on the manager's part. In other words, because the manager bears all the risk in the outcome of the risky project, the owner cares only about relative project cost and not the actual outcome of the risky project (if it is selected). In contrast, if the contract entails risk sharing, the information in realized cost is useful. Then, because only the realized cost of the risky project is informative, search and project selection incentives must be provided through the reward structure for the *risky* project. Specifically, search incentives can be provided only through the alternate to the default project. If the default is the risk-free project, search incentives must be provided through rewards for the risky project. Similarly, if the default is the risky project, search incentives must be provided through rewards (bounded by zero) for choosing the risk-free project and through penalties for failure of the risky project.

However, the default project depends on manager type; beliefs about *ex ante* profitability vary in manager type. Hence, search incentives must be provided through different routes for the different manager types. The optimal contract must partition the class 2 manager types into two subclasses that correspond to the manager's default project. Moreover, the same contract must be offered to all types within a subclass. This claim obtains from the truth-telling constraints on reporting of the search result and manager type. First, because search leads to private information, the payment for the risk-free project and the gamble in the outcome of the risky project must be invariant in the reported search result. Second, pre-contract asymmetry on manager type implies that the same contract must be offered to all manager types within a (sub)class.

The feature that all manager types within a subclass are offered the same contract facilitates characterizing the owner's optimal response as distortions in the information-acquisition and resource-allocation decisions. In particular, the payment scheme implicitly defines (and is defined by) the chosen cutoffs for manager class and conditional investment. The equivalence results from the truth-telling constraints (see Proposition 3). Hence, examining the owner's choice of the cutoffs is sufficient to characterize search incentives to the manager and thus, the owner's expected cost.

For intuition, note that search incentives for the class 2A types are implicit in the choice of M_t and explicit in the choice of r_{as}. Given r_{as}, raising M_t increases the range of search results for which the risky project is chosen and, thus, increases search incentives. Alternatively, given M_t, lowering r_{as} increases the *ex ante* probability of choosing the risky project and, thus, explicitly enhances search incentives. Similar implicit and explicit incentives are available to the class 2B manager types as well (through lowering M_h and raising r_{bs}). Proposition 3 indicates that the owner's optimal response uses a combination of the explicit and implicit means to provide search incentives to the manager (as expected). The following subsections provide additional intuition.[14]

Information Acquisition

Consider the information-acquisition decision. The main result is that increases in managerial risk aversion cause a strict reduction in the set of manager types asked to acquire information. Consider the effect of increasing risk aversion on the owner's choice of M_l. First, to induce search, the contract must impose risk on the class 2 manager types. Because the manager is risk-averse, the payment scheme must provide a risk-premium to the manager. Second, for given risk aversion, the value to the owner due to search by the type M_l manager is zero. Then, because risk-premium is increasing managerial risk aversion, $dM_l/d\lambda > 0$.

Next, consider the choice of M_h. First, for the class 2B manager types, the default is the risky project. Hence, search incentives must be provided by values for j that insure the manager against the gamble in the outcome of the risky project. Given r_{bs}, the range of such search results is decreasing in manager type. Hence, for manager types in the neighborhood of M_h, the cost (in the form of a risk-premium) of inducing search is increasing in manager type. Second, for given risk aversion, the value to the owner due to search by the type M_h manager must be zero. Then, because the risk-premium is increasing managerial risk aversion, increases in managerial risk aversion cause a reduction in the higher end of the manager types instructed to search as well; $dM_h/d\lambda < 0$.

Thus, increases in managerial risk aversion cause a strict reduction in the set of manager types asked to search for additional information. Hence, if the manager were extremely risk-averse, the set of manager types asked to search would be empty. The owner then instructs that project selection be conditioned on the manager's *ex ante* information alone. Because this strategy imposes no risk on the manager, it is optimal. Imposing any risk at all on the manager would be "too expensive," and search can be induced only by imposing risk. In sum, the owner's optimal response to information asymmetry always distorts the information-acquisition decision, relative to the first-best solution.

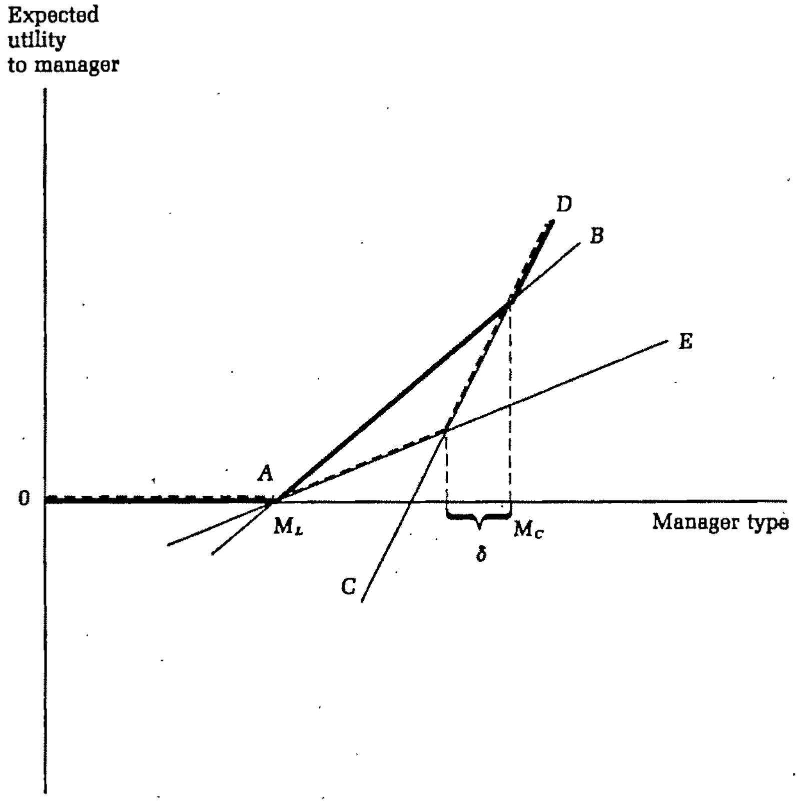
Resource Allocation

Consider the choice of the investment cutoff that defines the resource-allocation decision. The intuition here is similar to the intuition for why the manager class cutoffs deviate from the first-best solution. Specifically, search incentives are explicitly affected by the specified investment cutoff. The optimal contract trades the reduced cost of inducing search with the increase in project cost due to using other than the first-best

---

[14] Note that overinvestment in the risky project occurs for some manager types and for some states of nature. Hence, gains to renegotiation (after reporting of the search result) may exist. However, if renegotiation is allowed for, the manager has incentives to misrepresent private information because the renegotiation decision and the subsequent sharing of resulting gains will depend on the manager's messages and such dependence can be anticipated by the manager. (The revelation principle does not apply if the owner cannot credibly commit to the offered contracting scheme.) The assumption here is that the owner can credibly commit not to renegotiate. Derivation of a renegotiation-proof solution (if it exists) is outside the scope of this study. See Farrell and Maskin (1989) for a general analysis of renegotiation-proof solutions in repeated games.

Figure 2
Conditional Investment Decision



—— = Expected utility across manager types under contract with $r_m = r_f$.
--- = Expected utility across manager types under contract with $r_m < r_f$.

Note: For ease, a linear representation is used for expected utility.

AB = Expected utility from claiming class 2A status ($r_m = r_f$).
AE = Expected utility from modified contract ($r_m < r_f$).
CD = Expected utility from claiming class 2B status.

investment cutoff. The trade off leads to overinvestment in the risky project (in class 2A types) and to strict value for communication (in class 2B types).

Begin by noting that the choice of the risky project is perfectly informative as to whether the class 2A manager conducted the unobservable search. Unless the search result were favorable, the manager would not choose the risky project; the default choice is the risk-free project. Hence, for class 2A manager types, overinvesting in the risky project increases the probability that the project cost is informative about the manager's search activity. It follows that for any given M_t (and thus the optimal M_t), a gamble with a smaller spread is enough to motivate search (see eq. [A5] in the appendix). The savings inherent in a smaller risk-premium to the manager offset the cost of overinvestment.

Figure 2 provides additional intuition. Line AB represents the expected utility to the class 2A manager types if the contract does not overinvest (i.e., sets $r_{as}=r_f$). Line CD represents the expected utility from the contract offered to class 2B manager types. The two lines intersect at M_c as this manager type is indifferent (by definition) between claiming class 2A or 2B status. Further, line CD intersects line AB from below; otherwise, the truth-telling constraints would be violated. Next, let line AE represent the expected utility under a modified contract that sets r_{as} below r_f. Line AE is everywhere lower than line AB. Overinvesting in the contract offered to the class 2A types results in a modified contract with a smaller spread in the gamble on the cost of the risky project. However, because the contract to the class 2B manager types was not altered, the above modification also lowers M_c by δ. Then, as argued above, for manager types M_l to $M_c-\delta$, the owner benefits by overinvestment. Further, the expected payout for manager types $i>M_c-\delta$ is reduced because the contract to manager types $>M_c$ is not altered and the expected utility to manager types $M_c-\delta<i<M_c$ is lower.[15] Overinvesting in the risky project strictly benefits the owner. Contracting with a risk-averse manager distorts the resource allocation decision, relative to the first-best solution.

Examining the effect of r_{as} on search incentives established overinvestment in the risky project. Similar reasoning (the effect of r_{bs} on search incentives to the class 2B types) is used to establish strict value to communication. In particular, the default is the risky project for the class 2B manager types. For such manager types, underinvesting in the risky project allows more room to provide incentives. Underinvestment enlarges the *ex ante* probability that the alternate (the risk-free project) to the default project (the risky project) is chosen. Value to communication arises because, with communication, the incentive effect can be achieved by modification of the contract without actual underinvestment (see Proposition 1).[16] Together with Proposition 2, the result implies that regardless of the manager's risk preference, establishing communication channels (e.g., a bottom-up budgeting system) is strictly valuable.

III. Conclusion

Recent research argues that seemingly suboptimal practices in allocating resources to projects are in fact optimal responses to the information asymmetry between members of a firm. This article extends the inquiry by examining the generic resource allocation problem of choosing between two competing projects. In the stylized owner-manager firm, the owner delegates project choice because the manager is better able to acquire costly additional information and has private information on the cost-benefit tradeoff to such information search. In this situation, the owner cannot decompose the information-acquisition and the resource-allocation problems: by observing project choice alone, the owner cannot tell whether, in fact, the manager expended effort to search. Hence, the optimal contract must use the information in the realized cost (1) to motivate a subset of manager types to select after expending effort to search, and (2) to motivate the complementary set to choose the project based on their pre-contract information alone.

---

[15] Lowering r_{as} (and the attendant changes in $I2A$) also lowers M_c. That is, incentives to renege and implement the risky project are enhanced for the high manager types and provide a cap on the extent of reduction in r_{as}. To illustrate, consider the extreme case when $r_{as}=0$ (always choose risky project). Then, the market constraint and (SR 1) together imply that $I2A_f=I2A_s=a$. But then all manager types can renege and implement the risky project to obtain utility $a>0$ which is not optimal.

[16] Note that r_{bs} cannot be increased arbitrarily because of the truth-telling constraints on manager type. Specifically, set $r_{bs}=1$. Then, because $I2B_{ij,m}$ is bounded by zero (else, the class 1 agent types would claim class 2B status), the truth-telling constraints for the class 2B agent types are violated.

In this setting, if the manager is risk-neutral, the optimal contract distorts information acquisition, relative to the first-best solution. The distortion occurs because the optimal contract exploits the interaction between the manager's pre- and post-contract information to discriminate among the manager types. Such discrimination shifts to the right the set of manager types asked to search. The shift implies that some manager types are asked to search even though in the first-best solution their value to search is negative; other manager types are asked to forego search even though their first-best expected value to search is positive. Overall, the optimal contract distorts information acquisition in such a way that the risk-free project will be chosen more often than it would have been chosen in the first-best solution. Absent communication, this distortion also results in conditional underinvestment in the risky project. Reporting of the search result has strict value because it helps provide the same incentives as if the optimal contract called for conditional underinvestment.

In contrast, if the optimal contract entails risk sharing, both the information-acquisition and the resource-allocation decisions deviate from those in the first-best solution. The reason is that search incentives can be provided through distortions both in the search and the conditional investment decisions. In particular, increases in managerial risk aversion cause a strict reduction in the set of manager types instructed to search. Moreover, conditional overinvestment in the risky project always occurs. As in the case of a risk-neutral manager, however, reporting of the search result has strict value.

The results suggest that practices that seem to disregard the value of information search and that seem to use acquired information sub-optimally may be rational responses to the problem of delegating information acquisition and project selection to managers. Further, the results show benefit to establishing communication channels from the manager to the owner of the firm. Thus, the results lend support to the use of accounting data (e.g., bottom-up budgets) in managerial performance evaluation.

Appendix

Proof of Proposition 1

Assume to the contrary that the optimal contract to the type *i* manager satisfies $r_{iu}>r_{if}$. For the range of *j* values leading to $r_{if}\leq P_{ij}<r_{iu}$, add a contract that sets $I2_{iju}=I2_{ijf}=I2_{ijm}$, where $I2_{ijm}$ is the payment for the risk-free project in the assumed contract. Adding the additional contract does not change the search incentives to the manager in any way. Also, for $r_{if}<P_{ij}<r_{iu}$, the manager is now indifferent among the two projects. By construction, under the modified contract, the expected payments by the owner do not increase. However, the owner avoids the higher cost of the risk-free project for $r_{iu}>P_{ij}>r_{if}$. Contradiction.■

Proof of Proposition 2

The proof consists of several lemmas.

Lemma 2.1: In the optimal contract, the owner discriminates among manager types.

Proof: Consider a contract that sells the firm to the manager for price $V(M_i)>0$, where M_i is the lowest manager type willing to pay the price. Because $V(i\leq M_i^*)=0$, consider $M_i>M_i^*$.

Increase $V(M_\iota)$ by $\epsilon > 0$ and rebate ϵ_1 (satisfying $V(M_\iota) + \epsilon > \epsilon_1 > \epsilon$) for all $P_{ij} < r' < 1$ such that:

$$\Sigma_k C_k r' = C_m - \epsilon_1, \text{ and} \tag{A1}$$

$$\int \epsilon_1 P(j) dj = \epsilon \qquad \text{for } j: 0 \leq P_{ij}(i = M_\iota) < r'. \tag{A2}$$

That is, the incremental price is compensated via a rebate for the risk-free project. By construction, the modified contract yields the same expected utility to manager type M_ι. Under the modified contract, manager type M_ι would search (because $M_\iota > M_\iota^*$ implies value of search is positive). Under the modified contract, the expected gain to the owner is:

$$\int \epsilon \phi(i) di \qquad \text{for } M_\iota \leq i \leq i_{max}. \tag{A3}$$

If manager types $M_\iota \leq i \leq M_k$ subsequently search, the expected loss is:

$$\begin{array}{ll} \int_i \{ \int_j \epsilon_1 P(j) dj \} \phi(i) di & \textit{for } j: 0 \leq P_{ij} < r', \text{ and} \\ & \text{for } M_\iota \leq i \leq M_k \\ 0 \quad \text{for } M_k < i \leq i_{max}. & \end{array} \tag{A3$'$}$$

Note that $\int \epsilon_1 P(j)$ for $0 \leq P_{ij} < r'$ is decreasing in i due to interaction between manager type and search result. Hence, the net gain to the owner is (weakly) increasing in manager type. The claimed result follows because the net gain is weakly positive for all manager types and is strictly positive for at least one manager with type $M_\iota < i \leq i_{max}$ (because of the ordering of the manager types by increasing skill).■

Lemma 2.2: Absent communication, a contract that discriminates among manager types by providing a rebate for the risk-free project is optimal.

Proof: Because the manager is risk-neutral and because realized cost is the only observed variable, the owner need be concerned only with the relative expected cost for the two projects. Holding the cost of the risk-free project at C_m, and imposing an additional charge for the risky project is economically equivalent to providing a rebate for the risk-free project.

Holding the cost of the risk-free project at C_m and providing a rebate for the risky project is not optimal. Doing so causes the expected rebate to increase weakly in manager type. The *ex ante* probability of selecting the risky project is weakly increasing in manager type. Hence, such action is not optimal. Relative to selling the firm to the manager, such action reduces the information-rent to the owner because extracted information-rent is a function of the lowest manager type "renting" the firm.■

Lemma 2.3: In the optimal contract, $M_\iota > M_\iota^*$.

Proof: The modification in Lemma 2.1 is not feasible for $i \leq M_\iota^*$. Else, because $V(i \leq M_\iota^*) = 0$, manager types $i \leq M_\iota^*$ obtain utility $\epsilon_1 - \epsilon > 0$, which is not optimal. The owner therefore trades off the opportunity loss for manager types $M_\iota^* \leq i < M_\iota$ with the increased price obtained from types $i > M_\iota$. This tradeoff leads to $M_\iota > M_\iota^*$ (see also Antle and Eppen 1985).■

Lemma 2.4: In the optimal contract, reporting of the search result has value.

Proof: Absent communication, the optimal contract provides a rebate for the risk-free project for all $j: P_{ij} < r'$. Truth telling of search result implies $r' > r_f$. The contract entails underinvestment in the risky project. Communication allows for the modification without distorting the conditional investment decision (see Proposition 1). Hence, it has strict value.■

Lemma 2.5: In the optimal contract, $M_h > M_h^*$.

Proof: Consider the nontrivial case when $r' \leq M_h^*$. For manager type M_h^*, for $P_U \leq r'$, the rebate lowers the cost of the risk-free project. Hence, relative to the first-best solution, the rebate increases this manager type's opportunity cost (expected cost saving from search) of directly implementing the risky project; $r' > r_f$. Then, because the first-best expected cost saving to manager type $M_h^* = a$, sequential rationality implies $M_h > M_h^*$.■

Proof of Proposition 3

The proof consists of several lemmas.

Lemma 3.1: The optimal contract must partition the class 2 manager types into two subclasses. Further, the same contract is offered to all manager types within the same (sub)class.

Proof: In the described model, because $r_u \leq r_f$ (Proposition 1) and $r_f < M_h^*$ (see Balakrishnan 1989), at most one of the (IS) constraints would be binding for any given class 2 manager type. That is, the default project differs in manager type. Hence, incentives to search must be provided through differing channels (i.e., through results that lead to selection of the alternate to the default project). The optimal contract must partition the class 2 manager types into two subclasses as per the manager type's default project.

Proof of the second claim that the same contract is offered to all manager types within the same (sub)class is tedious but straightforward (see Balakrishnan 1989). The drivers for the result are the truth-telling constraints in manager type and search result and the feature that the default project is the same for all manager types within a (sub)class.■

Remark 1: Denote the set of class 2 manager types for whom the default is the risk-free (risky) project as class 2A (class 2B) manager types. The instructed investment cutoff is respectively denoted r_{as} and r_{bs}. Let M_c denote the manager type who defines the cutoff between the two subclasses. That is, manager type M_c is indifferent to classification either in class 2A or class 2B.

Remark 2: For truthful reporting of the search result, the same gamble on the outcome of the risky (risk-free) project must be offered for all reported search results. Else, the manager has incentives to misrepresent the observed search result.

Remark 3: Given Lemma 3.1 and Remark 2, Claim 1 is straightforward. For the class 2A types, the default is the risk-free project. Then, the (IS 2) constraint is binding for manager type M_c. The *ex ante* probability of choosing the risky project is increasing in manager type. Next, the (SR 1) constraint must be binding at $P_U = r_{as}$ for truthful reporting of the search result. Finally, the payment for the risk-free project must be zero. The class 1 manager types are paid zero and incentives to class 2A manager types are provided through rewards for implementing the risky project. Hence, the optimal payment scheme solves two equations in two unknowns.

Remark 4: The (TT) constraint on manager type M_c must be binding. (Expected utility is weakly increasing in manager type.) Next, the (SR 1) constraint is binding at $P_U = r_{bs}$ for truthful reporting of the search result (see Remark 2). Third, the (IS 1) constraint must be binding for manager type M_h. The *ex ante* probability that the risk-free project will be chosen is decreasing in manager type, and search incentives are provided through the alternate for the default project (see Balakrishnan 1989 for formal proof). Hence, the optimal payment scheme solves a system of three equations in three unknowns.

Remark 5: Due to the risk-averse nature of the manager, the same gamble on the outcome of the risky project is offered to manager types in class 2B and in class 3. Also, all the class 1 manager types are offered the same contract with $I1_{im}=0$. They choose the risk-free project without search.

Lemma 3.2: If the set of class 2A manager types is not empty, the optimal contract induces overinvestment in the risky project.

The following result is useful in proving the lemma.

Claim 3.2.1: If the set of class 2B and class 3 manager types is empty, the optimal contract induces overinvestment in the risky project.

Proof: Assume to the contrary that the optimal contract sets $r_{as}=r_f$. Substituting (SR 1) constraint into (IS 2) constraint and noting that $I2A_{ijm}=0$ yields: for $r_{as}<r\leq 1$ and for manager type M_t,

$$\int[U(I2A_{ijs})-U(I2A_{ijf})](r-r_{as})p(r)\,dr=a. \tag{A4}$$

Differentiating with respect to r_{as} and simplifying yields:

$$d[U(I2A_{ijs})-U(I2A_{ijf})]/dr_{as}>0. \tag{A5}$$

That is, the higher the instructed cutoff, the greater is the spread between the payments for the success or failure of the risky project.

Lower r_{as} below r_f. Then, for manager type M_t, the following costs and benefits accrue to the owner. First, higher expected project cost is incurred for all j such that $r_{as}<P_{ij}<r_f$. Second, because the manager is risk-averse, the expected payout by the owner for j such that $r_{as}<P_{ij}\leq r_f$ is higher. The benefit arises in that the information-rent for j such that $r_f<P_{ij}$ is reduced. Lowering r_{as} increases the region of j for providing search incentives, and the result is a smaller spread in the gamble offered to the manager (see equation [A5]). However, the probability of obtaining j such that $r_{as}<P_{ij}\leq r_f$ [$r_f<P_{ij}$] is decreasing (increasing) in the manager type. Thus, the expected cost (benefit) to the owner is decreasing (increasing) in manager type.

Finally, if $r_{as}=r_f$ characterizes the optimal contract, at the margin, the net benefit to the owner from lowering r_{as} is non-negative, for manager type M_t. The claim obtains because if the manager bears risk only for some $r_{as}>r_f$, the net benefit to the owner is strictly positive. It follows that the incremental costs would outweigh the incremental benefits for all the class 2A managers considered together. Contradiction.■

Proof of Lemma 3.2: Assume to the contrary that the optimal contract sets $r_{as}=r_f$ for the class 2A manager types. Modify the contract to lower r_{as} below r_f. Obviously, the modification would lower M_c by $\delta>0$. For manager types $i>M_c-\delta$, offer the same contract as originally offered to the class 2B types. It is easy to verify that the above modification preserves truth-telling (see figure 2). However, for manager types $M_t\leq i<M_c-\delta$ following claim 3.2.1, the above modification benefits the owner. Further, the expected utility to manager types $M_c-\delta<i<M_c$ is lower, *ceteris paribus*. Contradiction.■

Lemma 3.3: If the set of class 2B manager types is not empty, there is strict value to communication of the search result.

Proof (sketch): Because the default project for the class 2B manager types is the risky project, there is no benefit to setting $r_{b_s} < r_f$. Relative to setting $r_{b_s} = r_f$, setting $r_{b_s} > r_f$ has the benefit of increasing the region of j used to provide search incentives. Thus (following arguments analogous to the proof of claim 3.2.1), such action reduces the expected payout by the owner. From Proposition 1, r_{b_s} can be set above r_f without underinvesting in the risky project. But the state-independent transfer (for $r_f < i \leq r_{b_s}$) is feasible only if the manager can communicate the search result.■

Lemma 3.4: As managerial risk aversion increases, there is a reduction in the higher end of the class 2 manager types. That is, $dM_k/d\lambda < 0$.

Proof: Because the (IS 1) constraint is binding for manager type M_k, if managerial risk aversion increases, the reward for the risk-free project must be increased (or if it is at its upper bound of zero, the spread in the gamble for the risky project must be increased). For the class 2B manager types, search incentives are provided through rewards for the risk-free project or through penalties for failure of the risky project.

The required change in the reward structure increases the cost of inducing manager type M_k to search. Further, the value of search (to the owner) by manager type M_k is zero. Hence, $dM_k/d\lambda < 0$. For given risk aversion, (1) in the neighborhood of M_k, the cost of inducing search is increasing in i because probability of choosing the risk-free project is decreasing in i, and (2) the optimal contract has zero value (to the owner) due to search by manager type M_k.■

Remark 6: The change in the reward structure and increased risk aversion both influence r_{b_s} upward. Thus, underinvestment in the risky project could result due to the change in the reward structure; the resultant higher expected project cost to the owner may have to be considered. But with communication of the search result, incentive effects can be provided without underinvestment (Proposition 1). Hence, the change in the reward structure can be accomplished without altering the expected project cost to the owner.

Lemma 3.5: As managerial risk aversion increases, there is a reduction in the lower end of the class 2 manager types. That is, $dM_\iota/d\lambda > 0$.

Proof: The value of search (to the owner) by manager type M_ι is zero. The contract to manager type M_ι compensates him for the risk in conducting the search and in the outcome of the risky project (if it is selected). Then, because risk-premium increases in managerial risk aversion, $dM_\iota/d\lambda > 0$. The probability of obtaining $P_{ij} > r_{a_s}$ is increasing in manager type.■

Remark 7: The careful reader will note that the proof of Lemma 3.5 ignores one potential source of benefit to the owner. Specifically, while increased risk aversion would influence r_{a_s} upward, the increase in $I2A_u$ influences r_{a_s} downward. If the first effect dominates, because $r_{a_s} < r_f$, the change in the reward structure would result in less overinvestment in the risky project and thus lower expected project cost to the owner. This source for the benefit increases value of search (to the owner) and thus has a downward influence on M_ι. However, $dr_{a_s}/d\lambda \leq 0$. The benefit to overinvestment is from the reduction in risk-premium to the class 2A types and risk-premium increases in managerial risk aversion. Hence, in examining $dM_\iota/d\lambda$, the potential for benefits through reduced project cost is absent and is therefore ignored.

References

Antle, R., and G. Eppen. 1985. Capital rationing and organizational slack in capital budgeting. *Management Science* 31 (February): 163–74.

———, and J. C. Fellingham. 1990. Resource rationing and organizational slack in a two-period model. *Journal of Accounting Research* 28 (Spring): 1–24.

Balakrishnan, R. 1989. Information acquisition and resource allocation decisions. Working paper, University of Iowa, Iowa City.

———. 1990. Resource rationing with multiple divisions. Working paper, University of Iowa, Iowa City.

Baron, D., and R. Myerson. 1982. Regulating a monopolist with unknown costs. *Econometrica* 50 (July): 911–30.

Conroy, R., and J. Hughes. 1987. Delegating information gathering decisions. *The Accounting Review* 62 (January): 50–66.

Demski, J. S., and D. E. M. Sappington. 1984. Optimal incentive schemes with multiple managers. *Journal of Economic Theory* 33 (June): 152–71.

———, and ———. 1987. Delegated expertise. *Journal of Accounting Research* 25 (Spring): 68–89.

Farrell, J., and E. Maskin. 1989. Renegotiation in repeated games. *Games and Economic Behavior* 1 (December): 327–60.

Harris, C., H. Kriebel, and A. Raviv. 1982. Asymmetric information, incentives, and intra-firm resource allocation. *Management Science* 28 (June): 604–20.

Horngren, C. T., and G. Foster. 1987. *Cost Accounting: A Managerial Emphasis*. 6th ed. Englewood Cliffs, NJ: Prentice-Hall.

Jensen, R. 1982. Adoption and diffusion of an innovation of uncertain profitability. *Journal of Economic Theory* 27 (June): 182–93.

Lambert, R. A. 1986. Executive effort and selection of risky projects. *Rand Journal of Economics* 17 (Spring): 77–88.

Laffont, J., and J. Tirole. 1986. Using cost observations to regulate firms. *Journal of Political Economy* 94 (June): 614–41.

Melumad, N., and S. Reichelstein. 1987. Centralization vs. delegation and the value of communication. *Journal of Accounting Research* 25 (Supplement): 1–18.

———, and ———. 1989. Value of communication in agencies. *Journal of Economic Theory* 47 (April): 334–68.

Mussa, M., and S. Rosen. 1978. Monopoly and product quality. *Journal of Economic Theory* 18 (August): 301–17.

Myerson, R. 1982. Optimal coordination mechanisms in generalized principal-agent models. *Journal of Mathematical Economics* 10 (June): 67–81.

SHORTER ARTICLES

1. Robert M. Halperin and Bin Srinidhi
   U.S. Income Tax Transfer-Pricing Rules and Resource Allocation: The Case of Decentralized Multinational Firms
   (Submitted: August 1988. Accepted: June 1990.)

2. Shiing-Wu Wang
   The Relation Between Firm Size and Effective Tax Rates: A Test of Firms' Political Success
   (Submitted: February 1989. Accepted: September 1990.)

3. Mary Stone and John Rasp
   Tradeoffs in the Choice Between Logit and OLS for Accounting Choice Studies
   (Submitted: August 1988. Accepted: September 1990.)

Editor's Note: A "Shorter Article" is judged by the degree and recency of contribution.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 141–157

U.S. Income Tax Transfer-Pricing Rules and Resource Allocation: The Case of Decentralized Multinational Firms

Robert M. Halperin
Fordham University
Bin Srinidhi
Rutgers University

SYNOPSIS: This study examines how transfer pricing, in the presence of differential taxation, affects the resource allocation and profitability of a decentralized multinational enterprise (MNE) that uses the same transfer price for tax and performance evaluation purposes.

An analytical model is built in which a foreign manufacturing division transfers a single product to a U.S. distribution division, which, in turn, sells it in the marketplace as part of a final product. It is assumed that both divisions have complete information concerning all cost and revenue functions and that tax rates are higher in the United States than abroad. Given this assumption, it is shown that it is in the interest of both divisions to cooperate with each other, and a cooperative equilibrium is used throughout the analysis.

When differential tax rates exist, but there are no transfer-pricing rules imposed by the taxing authorities, it is shown that the MNE's optimal resource allocation is the same as in the absence of taxes; however, firm-wide profits are not maximized.

When the resale-price method of computing transfer prices (Reg. §1.482-2(e)(3)) is used, the results differ, depending on the manufacturing

We are grateful for the helpful comments received from participants in workshops at New York University, Columbia University, Indiana University, and the University of Waterloo on earlier drafts of this article. We would especially like to thank Rick Young and George Carlson, who discussed this study at the 1989 Tax Research Symposium held at the University of Illinois at Urbana-Champaign, as well as Josh Ronen, Kashi Balachandran, and the anonymous referees for their insightful comments.

Submitted August 1988.
Revisions received June and November 1989, and March 1990.
Accepted June 1990.

division's bargaining power. At low levels of bargaining power, the distribution division will "guarantee" a certain level of profit to the manufacturing division, and firmwide after-tax profits will be maximized. As the manufacturing division's bargaining power increases beyond a certain point, however, firmwide optimal profits will no longer be achieved. In addition, the final product and the most similar product (as defined by Reg. §1.482-2(e)(3)) will be produced beyond the firmwide after-tax optimum.

When transfer prices are computed in accordance with the cost-plus method (Reg. §1.482-2(e)(4)), it is shown that production of the final product first decreases, then increases, and production of the most similar product increases as the distribution division's bargaining strength increases. In addition, there is no guarantee of profit to the manufacturing division. Accordingly, it is unlikely that firmwide after-tax profits will be maximized. Demonstration of these results is achieved with the use of numerical examples.

The study concludes with a discussion of the tax policy implications of these results.

Key Words: *Multinational firms, Transfer pricing, Tax, Decentralization.*

IN Halperin and Srinidhi (1987) we examined the resource allocation effects in a centralized multinational enterprise (MNE) that is subject to tax transfer-pricing regulations under Federal Income Tax Regulations §1.482–2(e). Such an MNE trades off tax savings, realized by shifting income to the lower taxed country, against the opportunity cost of decreased profits caused by deviating from the no-tax optimum production decision. Such a trade off is shown to result in a resource allocation different from what would prevail in the absence of tax transfer-pricing regulations.

In this article, we extend our results to a decentralized MNE in which transfer pricing also is used for divisional performance evaluation. The divisions are autonomous, have different negotiating abilities, and can choose to cooperate or not with each other.

To justify an examination of this setting, we first must address two issues: (1) whether divisional profits computed using transfer prices are used to evaluate the performance of divisional managers, and (2) whether the same transfer-pricing scheme is used for tax and performance evaluation purposes. Tang (1980) surveyed 192 Canadian firms and found a majority cited performance evaluation as a dominant objective of transfer pricing. The issue of whether transfer prices are used predominantly for performance evaluation or for tax purposes or both was addressed in a survey by Wu and Sharp (1979). While domestic firms in their survey cited performance evaluation as the dominant reason for transfer pricing, MNEs cited compliance with tax regulations as most important. This includes compliance with foreign transfer-pricing rules, since foreign taxing authorities frequently have different rules from those in the United States. Though empirical evidence on MNEs' transfer-pricing motivations is not very extensive, the existing evidence suggests that divisional performance evaluation is at least one important reason for transfer pricing. According to a survey of 52 MNEs reported in Yunker (1982), some MNEs that give primary importance in transfer pricing to corporate profit maximization do not use divisional profits solely for performance evaluation purposes. This seems to be the only available study on the interrelationship between transfer pricing and performance evaluation criteria in MNEs and provides no evidence on *how* they modify divisional profits. In view of the current state of empirical research on this subject, we assume that tax transfer price-based divisional profits constitute at least one of the important performance evaluation criteria. If evidence becomes available in the future on how divisional profits are modified for performance evaluation, it would be interesting to see, analytically, whether these modifications ameliorate or exacerbate the resource allocation difference derived in this study.

On the second issue of whether MNEs use the same transfer pricing for both tax and internal purposes, Czechowicz et al. (1982) report that 89 percent of U.S.-based MNEs use the same transfer-pricing method for both purposes despite a potential resource misallocation from doing so. The reason for this is explained in the following quotation from a manager of a U.S. MNE:

> We've talked about developing a system that would (use) dual (transfer) prices—one for the tax authorities and one for internal control. But the cost and time needed to set up and administer such a system would be substantial. Furthermore, if the tax authorities find out that we are using a dual pricing system, we might encounter some difficulties. So we probably won't make any modifications in our system (Czechowicz et al. 1982, 61).

Therefore, for purposes of this study, we assume that decentralization and the divisional profit-based performance evaluation are prespecified. We do not examine the need for decentralization[1] or the possibility of having nonprofit-based compensation contracts for divisional managers.[2] We examine the resource allocation decisions of a decentralized U.S.-based MNE that transfers an intermediate good without an external market at the same transfer price for internal and tax-reporting purposes using transfer-pricing rules found in Reg. §1.482–2(e). We assume that the buying and selling divisions have complete information concerning the cost and revenue functions of the transferred product. Under this assumption, the two divisions will cooperate with each other and maximize firmwide profits if there are no restrictions on transfer prices. We call this maximum profit the *unconstrained optimum*. In Halperin and Srinidhi (1987) we showed that if transfer price restrictions are present, the centralized MNE's maximum profit will be less than the unconstrained optimum. We call this the *centralized optimum*. In this study, we show that:

1. Even if the divisions of a decentralized MNE do not face any transfer-pricing rules for tax purposes, but use the same transfer price for both tax and internal

---

[1] A purely economics-based need for decentralization has not been established. See Melumad et al. (1989) for an examination of this issue.

[2] One could speculate on a number of reasons as to why only profit-based compensation contracts are found in practice. One reason could be the ease of implementation and clear understandability. Also, if divisions deal with many products, it may not be practical to have different product-specific compensation contracts to evaluate the divisional manager. In this study, we do not consider how the tax rules and production decisions might affect the way compensation contracts are formulated in MNEs. We thank Rick Young for bringing this to our attention.

purposes, the MNE's total income is likely to be less than the unconstrained optimum. However, the MNE's resource allocation will be the same as under the unconstrained optimum.

2. If the divisions of a decentralized MNE are constrained to use §482 transfer-pricing rules for both tax and internal purposes, the MNE's total profit is likely to be smaller than the centralized optimum. Furthermore, the MNE's resource allocation will be different from that under the centralized optimum.

The difference between the unconstrained optimum and the centralized optimum (including the resource allocation differences) represents the social cost of the §482 transfer-pricing rules. The difference between the centralized optimum and the decentralized optimum represents the cost of decentralization under the tax environment.

The setting we use here is identical to the one used in Halperin and Srinidhi (1987) with the exception that the divisions are decentralized. The "resale-price method" and "cost-plus method" examined in both these studies are nonmarket transfer-pricing methods. The market-based comparable uncontrolled price method (CUP) is used by 24 percent of MNEs according to Benvignati (1985) and by 15–41 percent of the firms in studies quoted in the Study of Intercompany Pricing by the Treasury Department and Internal Revenue Service (1988; this study is also known as the "Section 482 White Paper"). Of the nonmarket methods, the resale-price and cost-plus methods are used predominantly by manufacturing MNEs, such as the one examined here. An explanation of these methods and how they can be modeled is found in Halperin and Srinidhi (1987).

The remainder of the article is organized as follows. In section I, we present the general decentralized model. Sections II and III present the models and the results of the analysis under the resale-price and cost-plus methods, respectively. Concluding remarks are presented in section IV.

I. The Model—No Transfer-Pricing Rule

Consider an MNE that has two divisions, a manufacturing (M) division located in a foreign country and a distribution (D) division located in the United States. Formally, M is a wholly owned subsidiary of D incorporated under the laws of the country in which it is located. Technically, M is a controlled foreign corporation pursuant to Internal Revenue Code (IRC) §957(a). We assume that M operates exclusively in the foreign country. No U.S. taxes are payable on M's income until it is paid to D as dividends (Reg. §1.954–3(a)(2)). We assume that M does not pay any dividends to D. M manufactures an intermediate product that does not have an external market and transfers it to D at a transfer price of r_t per unit, which is used for both tax reporting and performance evaluation purposes. D transforms this intermediate product into a final product. For efficient production and distribution, the quantity of intermediate product produced, q, should be equal to the quantity purchased by D. For ease of presentation, we assume that q also represents the quantity of the final product distributed by D. M and D pay income taxes at effective rates of t_m and t_d, respectively, and t_m is assumed to be less than t_d.

The cost of manufacturing the intermediate product, borne by M, is denoted by $C_m(q)$ and is characterized by a positive and increasing marginal manufacturing cost. D incurs distribution and transformation costs denoted by $C_d(q)$. Marginal distribution costs are also positive and increasing in q. The final product is sold in a market characterized by either perfect or imperfect competition. The revenue function, $R(q)$, is monotonically increasing and weakly concave in q.

We make the following additional explicit assumptions:

(1) both divisions have complete information on all cost and revenue functions;
(2) all agreements reached after negotiation between the two divisions are binding and enforceable; and
(3) the process of negotiation does not alter the preferences of the divisions.

Under these assumptions, it is always (weakly) Pareto-superior for the divisions to cooperate to maximize total firmwide profit and share it through bargaining. The bargaining process has been discussed by Dopuch and Drake (1964), Abdel-khalik and Lusk (1974), and Grabski (1985), among others. Although we do not explicitly model the process, we capture the results of bargaining by a number that gives the share of the total firmwide after-tax profit retained by D after negotiation.[3] We denote this number Θ, such that Θ lies between 0 and 1. Θ depends upon such factors as negotiation ability, size of the division, existence of local shareholders, and political and social pressures. D's bargaining strength is increasing in Θ. D and M have equal bargaining strength when $\Theta=0.5$.

In this section, we show that in the absence of any transfer-pricing rule other than the requirement that the same transfer price be used for both performance evaluation and tax purposes, the MNE's total after-tax profits will not be maximized. Necessary and sufficient conditions for firmwide after-tax profit maximization are:

1. $R'=C_d'+C_m'$ (where R', C_d', $C_m'=dR/dq$, dC_d/dq and dC_m/dq, marginal revenue and marginal costs of D and M, respectively) and
2. $r_t=(R-C_d)/q$ (i.e., all profits are taxed to M).

However, D will not cooperate with M to have all profits recorded and taxed to M.[4] By not cooperating with M, D can earn a profit of at least π_{d1}, which is the solution to:

$$\begin{aligned} &\underset{q,r_t}{\text{Maximize}} \quad T_m(r_t q-C_m) \\ &\text{subject to:} \quad q \epsilon \underset{\hat{q}}{\operatorname{argmax}}\, T_d(R-C_d-r_t\hat{q}), \end{aligned} \tag{1}$$

where $T_m=(1-t_m)$ and $T_d=(1-t_d)$.

Accordingly, firmwide after-tax profits are not maximized.

Notwithstanding the suboptimization of firmwide after-tax profits, we can show that the firm's resource allocation will be $R'=C_m'+C_d'$. By not cooperating with D, M can earn a profit of at least π_{m2}, which is the solution to:

---

[3] An alternative way of modeling would be to use a reservation utility for each division. Our model, in essence, uses a specific type of reservation utility based on divisional profits.

[4] The noncooperative game is examined here only to determine the lower bounds of the profits in the cooperative problem. The lower bound for D is obtained by allowing M to move first and assuming that D "reacts" to the transfer price. This sequence is important in determining the lower bound. Both divisions pick transfer prices and quantities only if there is a possibility of negotiation. Otherwise, the quantities and transfer prices will not match, and there will be no equilibrium.

$$\underset{q,r_t}{\text{Maximize}} \quad T_d(R-C_d-r_t q)$$
$$\text{subject to:} \quad q \in \underset{\hat{q}}{\operatorname{argmax}}\, T_m(r_t\hat{q}-C_m). \qquad (2)$$

The cooperative problem can be written as:

$$\underset{q,r_t}{\text{Maximize}} \quad T_d(R-C_d-r_t q)+T_m(r_t q-C_m)$$
$$\text{subject to (a):} \quad \frac{T_d(R-C_d-r_t q)}{T_d(R-C_d-r_t q)+T_m(r_t q-C_m)}=\Theta,$$
$$\text{(b):} \quad T_d(R-C_d-r_t q)\geq \pi_{d1}, \text{ and}$$
$$\text{(c):} \quad T_m(r_t q-C_m)\geq \pi_{m2}. \qquad (3)$$

Constraint (a) requires D's profit to be Θ percent of firmwide profits. Constraints (b) and (c) put lower bounds on the profits acceptable to D and M, respectively.

The solution to problem (3) is $R'=C_m'+C_d'$. Accordingly, the divisions will cooperate to insure that firmwide pretax profit is maximized and will then share this profit in accordance with their bargaining strengths, provided constraints (b) and (c) are satisfied.

This solution has an interesting property from a public policy perspective. It implies that taxes, per se, will not cause a decentralized MNE's optimal resource allocation to be different from that in the absence of taxation, even if tax rates differ across countries and transfer prices must be the same for tax and performance evaluation purposes. Thus, resource allocation distortions must be a result of transfer-pricing *methods* required by taxing authorities.

II. Resale-Price Method

The resale-price method is described in Reg. §1.482–2(e)(3), and its modeling is given in Halperin and Srinidhi (1987).[5] It requires the buying (reselling) division to identify one of its products that is most similar to the product bought from the selling division. The transfer price is computed by subtracting a markup from the selling price of the final product. The markup is the final product's selling price multiplied by the gross profit percentage (as a percentage of sales) earned by the buying (reselling) division on the most similar product.

Let $R_u(q_u)$ be the revenue function and $C_u(q_u)$ be the cost function of the most similar product, q_u. Following the usual economic assumptions, $R_u'>0$, $R_u''<0$, $C_u'>0$, and $C_u''>0$. For ease of exposition, we assume no cross-elasticity of demand between q and

---

[5] In Halperin and Srinidhi (1987) and in this article, the most similar product is identified within the MNE. Sometimes, it may become necessary to identify such a product outside the MNE. For example, if D is in the United States, a domestic U.S. firm producing and selling a product similar to the final product could be used as the yardstick to determine the gross profit percentage. However, typically, the MNEs are product leaders, and their production and sales divisions influence the gross profit percentage of the domestic firm. The resource allocation distortion examined in this study because of the transfer-price rule and decentralization could take place (the magnitude of distortion may be different) even when an outside firm is used. We are indebted to George Carlson of Arthur Andersen & Co. for pointing this out.

q_u. (Dropping this assumption will not change the nature of our results.) The transfer price r_t is computed as:

$$r_t = \frac{R}{q} - \frac{R}{q}\left(\frac{R_u - C_u}{R_u}\right) = \frac{R}{q}\frac{C_u}{R_u}. \tag{4}$$

Here too, we must solve the noncooperative equilibrium problems to obtain lower bounds on the profits acceptable to M and D. The minimum payoff that M gets by not cooperating is obtained when D dictates the transfer price to M and solves the following problem:

$$\begin{aligned} &\underset{q,q_u}{\text{Maximize}} \quad T_d(R - C_d - r_t q + R_u - C_u) \\ &\text{subject to: } q \in \underset{\hat{q}}{\operatorname{argmax}}\ T_m(r_t \hat{q} - C_m) \\ &\qquad r_t = \frac{R}{q}\frac{C_u}{R_u}. \end{aligned} \tag{5}$$

Denote the after-tax profit of M under problem (5) by π_{m5}. The minimum payoff that D gets by not cooperating can be expressed as the solution to the following problem when M dictates the transfer price to D.

$$\begin{aligned} &\underset{q,q_u}{\text{Maximize}} \quad T_m(r_t q - C_m) \\ &\text{subject to: } q, q_u \in \underset{\hat{q},\hat{q}_u}{\operatorname{argmax}}\ T_d(R - C_d - r_t \hat{q} + R_u(\hat{q}_u) - C_u(\hat{q}_u)) \\ &\qquad r_t = \frac{R}{q}\frac{C_u}{R_u}. \end{aligned} \tag{6}$$

Denote the after-tax profit of D under problem (6) by π_{d6}.

Under a cooperative equilibrium, the problem becomes:

$$\begin{aligned} &\underset{q,q_u}{\text{Maximize}} \quad T_d(R - C_d - r_t q + R_u - C_u) + T_m(r_t q - C_m) \\ &\text{subject to: } T_d(R - C_d - r_t q + R_u - C_m) \geq \\ &\qquad \Theta[T_d(R - C_d - r_t q + R_u - C_u) + T_m(r_t q - C_m)] \\ &\qquad r_t = \frac{R}{q}\frac{C_u}{R_u};\ T_m(r_t q - C_m) \geq \pi_{m5}; \\ &\qquad T_d(R - C_d - r_t q + R_u - C_u) \geq \pi_{d6}. \end{aligned} \tag{7}$$

Note that problem (7), unlike problem (3), contains an inequality in the profit-sharing constraint. In the absence of the resale-price constraint, the profit-sharing constraint is always binding, given the definition of Θ. However, in the presence of the resale-price constraint, this constraint is not necessarily binding. Assume that M has very low bargaining power. Under a noncooperative equilibrium (which represents M's minimum bargaining power), D would have to significantly underproduce q_u to reduce the transfer price and, consequently, to reduce the amount of q produced by M. Under a cooperative equilibrium, D can improve its own profit by producing more q_u and offering M a higher transfer price if M would voluntarily reduce q. In other words, it is in D's interest to share some profit with M because of the resale-price constraint. If M has very low bargaining power (i.e., if Θ is near 1) in a cooperative equilibrium, it might be in D's interest to offer more profit to M than that for which M can bargain. We can view the resale-price constraint as providing a nonbargained surplus to M. This explains why we need the inequality in the constraint. It also suggests that there is some region of Θ in which the constraint is not binding. If the constraint is not binding, the solution to problem (7) (found in the appendix, sec. A) is:

$$R' - C_d' - \frac{T_m}{T_d} C_m' = -R' \frac{C_u}{R_u}\left(\frac{T_m}{T_d} - 1\right); \tag{8}$$

$$R_u' + R\left(\frac{T_m}{T_d} - 1\right)\frac{d}{dq_u}\left(\frac{C_u}{R_u}\right) = C_u'. \tag{9}$$

The solution to simultaneous equations (8) and (9) gives a unique optimal value of q, denoted q_r^*, and a unique optimal value for q_u, denoted q_{ur}^*. Thus, D will offer a transfer price $r_r(q_r^*, q_{ur}^*) = (R(q_r^*)/q_r^*)(C_u(q_{ur}^*)/R_u(q_{ur}^*)$ per unit if M transfers q_r^* to D, and zero otherwise. M will accept the offer if the income it will receive is greater than or equal to π_{m6}.

A comparison of equation (8) with the condition $R' = C_d' + C_m'$ reveals the effect of the resale-price method with the nonbinding first constraint on the production of q. Since the constraint on profit sharing is not binding, the resource allocation obtained and the payoffs realized are the same as the centralized optimum of Halperin and Srinidhi (1987). It is not possible to generalize whether q_r^* will be higher or lower than the unconstrained optimal q. The most that can be said is that the optimal q is likely to be different. A comparison of equation (9) with the condition $R_u' = C_u'$ reveals the effect of the resale-price method with the nonbinding first constraint on the production of q_u. From equation (9), $R_u' < C_u'$, i.e., q_{ur}^* is more than the unconstrained optimal level of q_u.

As Θ decreases, M can bargain for more profit than D is induced to give voluntarily. The second constraint will then be binding and the solution (found in the appendix, sec. B) becomes:

$$\frac{R' - C_d' - C_m'}{\left(1 - \frac{C_u}{R_u}\right)R' - C_d'} - 1 = \frac{R\frac{d}{dq_u}\left(\frac{C_u}{R_u}\right)}{R_u' - C_u' - R\frac{d}{dq_u}\left(\frac{C_u}{R_u}\right)} = \frac{T_d}{T_m}\,\frac{(1 + \lambda - \lambda/\Theta)}{1 + \lambda}, \tag{10}$$

where λ is the Lagrangian multiplier for the first constraint.

For a given set of revenue and cost functions, the change in q and q_u associated with a change in Θ can be studied from equation (10).

In the example shown in table 1, the first constraint is not binding in the range $\Theta = (0.578, 1)$. In this region, $q_u = 228.08$, which is more than its unconstrained optimal level of 198.02. A comparison with the centralized optimum, also given in table 1,

Table 1

The Impact of Bargaining Power on the Resource Allocation Decisions and Profits: Resale-Price Method

	Θ	q	q_m	r_r	*PD*	*PDQ*	PDQ_m	*PM*	*Total Profit*
	0.578 to 1	309.76	228.08	259.63	58,179	38,833	19,346	42,497	100,676
	0.55	311.82	241.96	273.15	55,329	36,502	18,827	45,269	100,599
	0.50	314.63	263.06	296.87	50,130	32,437	17,666	50,102	100,205
	0.45	316.66	280.55	320.13	44,842	28,480	16,362	54,807	99,649
	0.40	318.17	295.40	343.09	39,616	24,603	16,013	59,424	99,040
	0.35	319.31	308.22	365.88	34,450	20,781	13,669	63,979	98,429
	0.30	320.16	319.45	388.57	29,354	16,999	12,355	68,493	97,847
Not Feasible	0.25	320.81	329.40	411.23	24,327	13,242	11,085	72,981	97,308
	0.20	321.30	338.30	433.92	19,365	9,500	9,865	77,457	96,821
	0.10	319.62	353.60	479.57	9,601	2,023	7,578	86,409	96,010
Centralized Optimum	0.578*	309.76	228.08	259.63	58,179	38,833	19,346	42,497	100,676
Noncooperative Equilibria	D^1	502	5	100.5	31,275	30,246	1,033	15,166	46,442
	M^1	157	327	503	24,949	18,553	11,396	45,998	5,947

Tax rates: t_d: 50% t_m: 40%

Table 1—*Continued*

Notation	*Description*	*The Cost and Revenue Functions Used*
Θ	D's share of total profit	$R=1000q-q^2$
q	Quantity of final product	$C_d=.5q^2+25q$
q_m	Quantity of most similar product	$C_m=0.1q^2$
r_t	Transfer price	$R_u=500q_m-q_m^2$
PD	D's total profit	$C_u=100q_m+0.01q_m^2$
PDQ	D's profit from q	
PDQ_m	D's profit from q_m	
PM	M's total profit	

* The Θ is computed for the centralized optimum as $PD/(PM+PD)$.

[1] The division with all bargaining strength.

reveals that both the resource mix and profit levels under decentralization for $\Theta > 0.578$ are indeed identical to the centralized optimum. From table 1, it can be seen that q and q_s increase when Θ decreases in the range of $\Theta = (0.578, 0)$. The intuition behind this result is that a decrease in Θ requires D to transfer a higher share of profits to M by increasing the transfer price. Under the resale-price constraint, increasing the transfer price is achieved only by producing q and q_s above centralized optimum levels. The total profits are at the centralized optimum for $\Theta \geq 0.578$. As expected, for $\Theta < 0.578$, M's after-tax profit increases, D's decreases, and the total after-tax profit of the firm decreases as Θ decreases. It may be noted that the range of $\Theta \epsilon [0.25, 0]$ is not feasible because D can get a profit of $24,949 by not cooperating. In the absence of the resale-price method constraint, D would have sacrificed a dollar of its after-tax profit to achieve a dollar increase in M's profit. The resale-price constraint forces D to sacrifice *more* than a dollar in after-tax profits to satisfy M's demand for every additional dollar in profits. From the noncooperative equilibrium solution, it is also clear that even when D has all the bargaining strength, it is in D's interest to cooperate with M and offer a higher transfer price for a smaller q, such that M's profit is increased. The interesting aspect of this result is that the resale-price method bestows on M a minimum bargaining strength.

In summary, when Θ decreases, D is forced to share more profits with M and adjust the levels of q and q_s to increase the transfer price within the resale-price framework. The output mix, and hence the extent of resource allocation distortion, is determined by the transfer-pricing rule in conjunction with the bargaining strengths of the divisions.

III. Cost-Plus Method

In this section, we assume that the intermediate product is transferred by M to D using the cost-plus method. Reg. §1.482–2(e)(4) requires the transfer price of the intermediate product to be its manufacturing cost plus a markup. The markup allowed is computed by multiplying the intermediate product's cost by the gross profit-to-cost ratio of a most similar product sold by M in an uncontrolled market. Let q_s represent the quantity of the most similar product sold by M, $R_s(q_s)$ the revenue, and $C_s(q_s)$ the costs associated with q_s. We assume that $R_s' > 0$, $R_s'' < 0$, $C_s' > 0$, and $C_s'' > 0$.

As in Halperin and Srinidhi (1987), we model the cost-plus method transfer price as:

$$r_t = \frac{C_m}{q}\left[1 + \frac{R_s - C_s}{C_{2s}}\right] = \frac{C_m}{q}\,\frac{R_s}{C_s}. \tag{11}$$

As in the resale-price method, we formulate the cooperative problem as follows:

$$\underset{q,q_s}{\text{Maximize}} \quad T_m(r_t q - C_m + R_s - C_s) + T_d(R - C_d - r_t q)$$

$$\text{subject to:} \quad T_d(R - C_d - r_t q) = \Theta[T_d(R - C_d - r_t q) + T_m(r_t q - C_m + R_s - C_s)]$$

$$r_t = \frac{R_s}{C_s}\,\frac{C_m}{q};$$

$$T_m(r_t q - C_m + R_s - C_s) \geq \pi_m; \text{ and}$$

$$T_d(R - C_d - r_t q) \geq \pi_d. \tag{12}$$

π_m and π_d are the lower bounds on M's and D's profits, respectively, obtained by solving the noncooperative problems with equation (11) as the constraint.

Solution to this problem, found in the appendix (sec. C), is the following:

$$\frac{T_m}{T_d}\left[1+\frac{R_s'-C_s'}{\dfrac{d}{dq_s}\left(\dfrac{R_s}{C_s}\right)C_m}\right]=-\frac{T_m\left(\dfrac{R_s}{C_s}-1\right)C_m'}{T_d\left(R'-C_d'-\dfrac{R_s}{C_s}C_m'\right)}=\frac{(1+\lambda)(1-\Theta)}{(1+\lambda)\Theta-1}. \tag{13}$$

Since the objective function in equation (12) is a monotonically increasing function of q, the profit-sharing constraint always holds as an equality.

A knowledge of the functional forms of the cost and revenue functions is required to determine the direction of movement of q and q_s with Θ. Typically, q_s increases with Θ because an increase in q_s depresses the transfer price, allowing D to get a greater proportion of the earnings. This is seen from table 2.

In the example shown in table 2, q_s increases with Θ, q decreases in the range (0, .7), but increases thereafter. The range (0, 0.1) of Θ is not feasible because D can get at least $12,765 by not cooperating, and the range (.8, 1) is not feasible because M can get at least $16,492 by not cooperating.

To transfer a greater share of the profit to D, the transfer price has to be decreased. This is achieved by overproducing q_s compared with the centralized optimum, thereby depressing the ratio (R_s/C_s). It can also be done by underproducing q and thereby decreasing the ratio (C_m/q). However, a reduction in q also would have the effect of decreasing the profit of D. This is why q is not monotone in Θ. It decreases over a part of the range, then starts increasing slowly.

It can be seen from table 2 that total after-tax profits reach the centralized optimum at $\Theta \cong 0.4$. (In fact, the value of Θ at which the centralized optimum is reached can be shown to be 0.394.) M's profit decreases and D's increases with Θ. Even though an increase in Θ causes more profit to be shifted to D, the profit of M from q_s increases over a range of Θ from 0 to 0.5. Still, the decrease in profit from q overwhelms the increased profit from q_s for M. The cost-plus method, in conjunction with the profit-sharing constraint, forces one division to sacrifice more than a dollar in after-tax profits so that the other division can make a dollar in profit.

From the above discussion, it is clear that the value of Θ is critical in determining the distortions in resource allocation when compared to the centralized optimum. There is always a value of Θ that yields the centralized optimum (say, Θ_{opt}). A knowledge of the specific cost and revenue functions is required to determine the quantities q and q_s. In the example, for $\Theta < \Theta_{\text{opt}}$, the most similar product is underproduced and the final product is overproduced in relation to the centralized optimum.

When Θ increases, M can share more profits with D only by making appropriate production decisions to decrease the transfer price within the cost-plus framework.

IV. Conclusions

In this study, we examined how transfer pricing, in the presence of differential taxation, affects the resource allocation and profitability of a decentralized MNE that uses the same transfer price for tax and performance evaluation purposes. We modeled the divisions of the MNE as maximizing their profits under a cooperative equilibrium with the condition that, after negotiation, the profit of any division would be no lower

Table 2
The Impact of Bargaining Power on the Resource Allocation Decisions and Profits: Cost-Plus Method

	Θ	q	q_s	r_s	*PM*	*PMQ*	PMQ_s	*PD*	*Total Profit*
Not Feasible	0	302.53	62.22	642.22	70,304	61,660	8,644	0	70,304
	0.1	292.12	71.34	604.70	64,270	54,788	9,482	7,141	71,411
	0.2	280.47	81.90	563.25	57,903	47,588	10,315	14,476	72,379
	0.3	267.04	94.49	517.02	51,168	40,052	11,115	21,929	73,097
	0.4	251.71	110.23	464.92	44,023	32,201	11,822	29,349	73,372
	0.5	235.12	131.50	406.48	36,431	24,174	12,257	36,431	72,862
	0.6	221.62	162.07	346.40	28,428	16,591	11,837	42,642	70,890
	0.7	218.37	198.43	299.55	20,366	10,636	9,730	47,522	67,888
Not Feasible	0.8	221.97	229.65	269.13	12,810	6,282	6,528	51,241	64,051
	0.9	227.11	253.83	248.06	6,020	2,854	3,166	54,180	60,200
Centralized Optimum	0.394*	252.67	109.18	468.17	44,456	32,671	11,785	28,917	73,373
Noncooperative Equilibria	D^1	135.65	0	338	16,492	16,492	0	33,065	49,557
	M^1	152.34	136.36	639.84	56,834	44,561	12,273	12,765	69,599

Tax rates: t_d: 50% t_m: 40%

Table 2—*Continued*

Notation	*Description*	*The Cost and Revenue Functions Used*
Θ	D's share of total profit	$R=1000q-q^2$
q	Quantity of final product	$C_d=0.1q^2+25q$
q_s	Quantity of most similar product	$C_m=q^2$
r_t	Transfer price	$R_s=500q_s-q_s^2$
PD	D's total profit	$C_s=200q_s+0.1q_s^2$
PMQ	M's profit from q	
PMQ_s	M's profit from q_s	
PM	M's total profit	

* The Θ is computed for the centralized optimum as $PD/(PM+PD)$.

[1] The division with all bargaining strength.

than the profit that division would have earned without cooperating. We showed that in the absence of any additional constraints on transfer-pricing policies, the resource allocation of the MNE will be the unconstrained optimum. However, the MNE's firmwide after-tax profits will not be maximized.

When transfer-pricing policies were limited to the resale-price or cost-plus methods, our results changed dramatically. We showed that the use of these methods for internal pricing distorts the resource allocation pattern both vis-à-vis the unconstrained optimum and the centralized optimum analyzed by Halperin and Srinidhi (1987).

Although general assertions could not be made about the production and allocation decisions, we showed that specific results can always be obtained for given cost and revenue functions. To demonstrate this, examples were solved under both methods to show how the profits and the production quantities differ from the centralized optimum and the unconstrained optimum.

In the example for the resale-price method, we showed that the negotiation constraint is not binding for low values of M division's negotiating power. In such a situation, the rule bestows upon M a minimum bargaining strength such that the decentralized optimum is achieved. Below a critical level of Θ, there is an effective bargained solution resulting in some range in which both the most similar and the final products are overproduced.

In the cost-plus method, the negotiation constraint is always binding. There is one value of Θ at which the decentralized optimum is the same as the centralized optimum. In the example, the similar product was underproduced (vis-à-vis the centralized optimum) below that level of Θ and was overproduced above that level. The final product was overproduced for low values of Θ. For a significant part of the range of Θ, the q and q_s values moved in opposite directions, unlike in the resale-price method.

These results demonstrate the types of distortions that can result in a decentralized organization under tax transfer-price rules. While the reduction in the optimal profit in the centralized case represents the joint cost of having differential taxes and transfer-pricing rules, this study examines the additional cost of decentralization. This incremental cost of decentralization is useful to the MNE management in efficiently trading it off with the political costs of modifying the transfer prices for internal purposes as suggested in Yunker (1982).

Much of the literature in taxation is devoted to understanding the direct effects of taxation on resource allocation through specific decisions, such as the labor-leisure choice or amounts of charitable contributions, or consumption of nontaxable perquisites. Recently, studies have begun to examine indirect effects of taxation on the choice of organizational forms, such as corporations and partnerships. These studies implicitly assume that different organizational arrangements affect resource allocation. If we take decentralization as a given, our study examines the direct effects of taxation on resource allocation. Our study can also be viewed as examining the indirect effect of taxation on the decision to centralize or decentralize.

Organizational forms are chosen, among other reasons, to achieve efficient risk sharing and provide proper incentives. Tax rules affect these choices, and it is of interest to policymakers to understand how tax policy affects organizational choice. From a policy perspective, there are two questions that are relevant to this study. The first is, "Can a transfer-pricing scheme be devised that will not affect the decision to decentralize or centralize?" The second is, "Given an organizational form, can a transfer-pricing scheme be devised that will not affect resource allocation (i.e., one that is consistent with the unconstrained optimum)?" In this study, we showed that the requirement that a decentralized MNE use the same transfer-pricing method for both tax and internal reporting purposes will not distort resource allocation. In Halperin and Srinidhi (1987) we showed that if a centralized MNE is required to use marginal-cost transfer pricing, the resulting resource allocation will not be distorted. Accordingly, the answer to the second question is "yes." The answer to the first question, however, is "no." It appears that any negotiated transfer price will very likely deviate from marginal cost.

Obviously, the major objective of the resale-price and cost-plus methods of transfer pricing is to allocate income fairly between taxing jurisdictions. We have shown that the use of either of these methods causes efficiency losses for decentralized MNEs.

Appendix

Section A. Solution to Problem (7), First Constraint Not Binding

Let π be the function to be maximized. Substituting the remaining constraint into the objective function gives:

$$\pi = T_d\left(R - C_d - R\frac{C_u}{R_u} + R_u - C_u\right) + T_m\left(R\frac{C_u}{R_u} - C_m\right). \tag{A1}$$

Taking the relevant partial derivatives, setting them equal to zero and rearranging gives:

$$R' - C_d' - \frac{T_m}{T_d}C_m' = R'\frac{C_u}{R_u}\left(1 - \frac{T_m}{T_d}\right); \tag{A2}$$

$$R_u' + R\left(\frac{T_m}{T_d} - 1\right)\frac{d}{dq_u}\left(\frac{C_u}{R_u}\right) = C_u'. \tag{A3}$$

Section B. Solution to Problem (7), First Constraint Binding

Problem (8) can be written as the Lagrangian function:

$$\begin{aligned}\mathcal{L} = {} & T_d\left(R - C_d - R\frac{C_u}{R_u} + R_u - C_u\right) + T_m\left(R\frac{C_u}{R_u} - C_m\right) \\ & + \lambda\left[\left\{T_m + \left(\frac{1-\theta}{\theta}\right)T_d\right\}R\frac{C_u}{R_u} - T_m C_m \right. \\ & \left. - \left(\frac{1-\theta}{\theta}\right)T_d\left(R - C_d + R_u - C_u\right)\right],\end{aligned} \tag{B1}$$

where λ is the Lagrangian multiplier. Taking the relevant partial derivatives and setting them equal to zero and simplifying gives:

$$\frac{R'-C_d'-C_m'}{\left(1-\frac{C_u}{R_u}\right)R'-C_d'}-1=\frac{R\frac{d}{dq_u}\left(\frac{C_u}{R_u}\right)}{R_u'-C_u'-R\frac{d}{dq_u}\left(\frac{C_u}{R_u}\right)}=\frac{T_d}{T_m}\,\frac{(1+\lambda-\lambda/\Theta)}{(1+\lambda)}. \tag{B2}$$

Section C. Solution to Problem (13)

Substituting for r_t from the cost-plus requirement and writing the Lagrangian function:

$$\mathcal{L}=T_m\left(\frac{R_s}{C_s}C_m-C_m+R_s-C_s\right)+T_d\left(R-C_d-\frac{R_s}{C_s}C_m\right)+\lambda\left[T_d\left(R-C_d-\frac{R_s}{C_s}C_m\right)\right.$$

$$\left.-\left(\frac{\Theta}{1-\Theta}\right)T_m\left(\frac{R_s}{C_s}C_m-C_m+R_s-C_s\right)\right]. \tag{C1}$$

Taking the partial derivatives and setting them equal to zero and simplifying gives:

$$\frac{T_m}{T_d}\left[1+\frac{R_s'-C_s'}{C_m\frac{d}{dq_s}\left(\frac{R_s}{C_s}\right)}\right]=-\frac{T_m\left(\frac{R_s}{C_s}-1\right)C_m'}{T_d\left(R'-C_d'-\frac{R_s}{C_s}C_m'\right)}=\frac{(1+\lambda)(1-\Theta)}{(1+\lambda)\Theta-1}. \tag{C2}$$

References

Abdel-khalik, A. R., and E. J. Lusk. 1974. Transfer pricing—A synthesis. *The Accounting Review* 49 (January): 8–23.

Benvignati, A. M. 1985. An empirical investigation of international transfer pricing by U.S. manufacturing firms. In *Multinationals and Transfer Pricing*, edited by A. M. Rugman and L. Eden. New York: St. Martin's Press, 193–211.

Czechowicz, I. J., F. D. S. Choi, and V. B. Bavishi. 1982. *Assessing Foreign Subsidiary Performance Systems and Practices of Leading Multinational Companies.* New York: Business International Corporation.

Dopuch, N., and D. H. F. Drake. 1964. Accounting implications of a mathematical programming approach to the transfer pricing problem. *Journal of Accounting Research* 2 (Spring): 10–24.

Grabski, S. V. 1985. Transfer pricing in complex organizations: A review and integration of recent empirical and analytical research. *Journal of Accounting Literature* 4: 33–71.

Halperin, R., and B. Srinidhi. 1987. The effects of the U.S. income tax regulations' transfer pricing rules on allocative efficiency. *The Accounting Review* 62 (October): 686–706.

Melumad, N., D. Mookherjee, and S. Reichelstein. 1989. Hierarchical decentralization of incentive contracts. Unpublished working paper, Stanford University.

Tang, R. Y. W. 1980. Canadian transfer pricing practices. *CA Magazine* 113 (March): 32–38.

Treasury Department and Internal Revenue Service. 1988. *Section 482 White Paper, A Study of Intercompany Pricing.* Washington, D.C.: Government Printing Office.

Wu, F. H., and D. Sharp. 1979. An empirical study of transfer pricing. *The International Journal of Accounting Education and Research* 12 (Spring): 71–99.

Yunker, P. J. 1982. *Transfer Pricing and Performance Evaluation in Multinational Corporations, A Survey Study.* New York: Praeger.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 158–169

The Relation Between Firm Size and Effective Tax Rates: A Test of Firms' Political Success

Shiing-wu Wang
University of Arizona

SYNOPSIS: Prior research has investigated firms' success in the political process and its relation to firm size. For example, effective tax rates have been used to measure the extent to which firms are able to influence corporate income taxation. Effective tax rates, however, are a biased proxy of political success because they may contain a factor, net operating losses (NOLs), attributable to firms' operating results and experience rather than to success in the political process. Moreover, if firm size is also correlated with NOLs, then NOLs may affect the relationship between firm size and tax rates in a systematic manner. If such an effect is substantial, the association of firm size and effective tax rates will be a biased basis for drawing inferences about firm size and political success. The purpose of this article is to assess the extent of the effect of NOLs on the overall relationship of firm size and tax rates. Path analysis is employed to isolate the effect of firm size on effective tax rates that is attributable to NOLs (i.e., the indirect effect of NOLs). Determining the indirect effect of NOLs allows an assessment of the bias in the overall firm size/tax rate relation attributable to NOLs.

To estimate the indirect effect of NOLs, a path model with two structural equations is built to identify the indirect links between firm size and effective tax rates mediated by NOLs, and the remaining direct link between firm size and effective tax rates. To correct errors-in-variables problems and to provide a consistent interpretation of the results, variables in the structural equations are modeled by a latent-variable approach (measurement model). The path model and the measurement model are esti-

This article is based on part of my doctoral dissertation. I am indebted to my dissertation chairman, William Kinney, and other dissertation committee members. I also gratefully acknowledge the suggestions of the anonymous referees.

Submitted February 1989.
Revisions received August 1989, and January, June, and September 1990.
Accepted September 1990.

mated by LISREL. Based on the parameter estimates of the structural equations, the path analysis results indicate that the indirect effect of NOLs has a statistically significant impact on the overall relationship between firm size and effective tax rates. This indicates that NOLs constitute an important omitted variable in prior studies. While effective tax rates may measure some degree of firms' success in the political process, conclusions regarding the relation between firm size and firms' success in the political process is confounded by the serious omission of NOLs.

Key Words: *Political success, Firm size, Effective tax rate, Path analysis.*

Data Availability: *Data used in this study will be available upon request to the author.*

A PARTIAL measure of success in the political process has been defined as a ratio of tax paid to some measure of income (see, e.g., Siegfried 1972; Zimmerman 1983), which is known as the effective tax rate.[1] Success in the political process is assumed to lead to special tax breaks, including excess deductions from taxable income, exclusions from taxable income, and tax credits. A firm with greater tax breaks reports lower taxable income and lower taxes paid for a given level of income. Consequently, effective tax rates computed on that basis would be low for firms with greater success in obtaining special tax breaks. That is, the variation in average effective tax rates measures the extent of firms' political success in influencing corporate income taxation.[2]

Effective tax rates may measure success in the political process with error, however, because they may contain a factor (NOLs) that is related to the firms' operating characteristics and experience rather than their success in the political process. The occurrence and utilization of NOLs normally eliminates the payment of taxes, resulting in low effective tax rates. To the extent that effective tax rates are affected by NOLs, they are a biased measure of firms' political success.

When NOLs are omitted as a variable in the analysis, the econometrics literature suggests that the estimated firm size/tax rate relation is biased and inconsistent if the omitted variable (NOLs) is correlated with the explanatory variable of firm size (Kmenta 1971). In other words, the overall firm size/tax rate relation may contain the confounding effect of NOLs if NOLs are omitted as a variable from the analysis and may be a biased test of the firm size/political success relation. The extent of the confounding effect of NOLs on the overall firm size/tax rate relation is assessed in this study by means of path analysis. The remainder of this article is organized as follows. Section I describes the model, sample composition, data sources, and variable definitions. Section II presents the results of the analyses. Section III contains a summary.

---

[1] Effective tax rates are a partial measure of firms' success in political process because they represent political success only in the area of corporate income taxation. Zimmerman (1983, 120) stated, "While all corporations face the same marginal federal tax rate schedule, numerous tax provisions like the investment tax credit and oil depletion allowance cause firms' effective tax rates to vary across firms and industries. This variation in effective tax rates provides a partial measure of firms' success in the political process. . . . Effective corporate tax rates are only a partial measure of the firms' political costs because tax rates exclude other political costs/benefits, such as antitrust, regulation, government subsidies and contracts, import quotas and tariffs."

[2] Firms are also concerned with the marginal income tax they pay and the variation in marginal effective tax rates may be a partial measure of firms' success in the political process. This study limits the discussion to average effective tax rates to assess the intervening role of NOLs on the overall firm size/average tax rate relation.

I. Research Design

Proposed Models

The extent of the effect of NOLs on the overall firm size/tax rate relation can be assessed by path analysis, which decomposes a given causal relation into its components. (An excellent review of path analysis can be found in Duncan 1975.) Path analysis allows for decomposition of the overall firm size/tax rate relation into two components: an effect of NOLs on the overall firm size/tax rate relation (the indirect effect of NOLs) and the remaining effect of firm size on tax rate. The first component indicates the extent of the overall firm size/tax rate relation attributable to NOLs.

First, a path model was developed to identify the direct and indirect links between firm size and the effective tax rate. It was postulated that small firms may be more likely to incur operating losses than large firms. Small firms are mostly new and growing companies that commonly incur NOLs because of high start-up costs and unsteady sources of revenues due to the lack of a broad client base. In addition, small firms are, in general, less diversified than large firms and are more sensitive to changes in economic conditions. In the event of a poor economy, narrowly focused small firms are more likely to be hurt by a decline in demand for goods or services and are more likely to incur NOLs than are large and mature firms. None of these factors is reflective of political success (or costs). Rather, they reflect the position of firms in the economy and the life cycle of the firm.[3] When firms incur NOLs, they normally pay no income tax, which reduces effective tax rates to close to zero. Thus, the path model is constructed around two causal links: firm size to NOLs and NOLs to tax rate. In addition, the direct causal link between firm size and effective tax rates suggested by prior research is retained to represent the remaining effect of firm size on effective tax rates.

These three causal relations result in two structural equations:

$$NOLS=\gamma_1 SIZE+\zeta_1, \tag{1}$$

$$ETR=\beta_1 NOLS+\gamma_2 SIZE+\zeta_2, \tag{2}$$

where $NOLS$ is net operating losses, $SIZE$ is firm size, ETR is the effective tax rate, γs and β_1 are parameter estimates, and ζs are errors in equations. The indirect effect of NOLs on the overall firm size/tax rate relation is then computed by the product of the γ_1 and β_1.

Success in the political process normally leads to future changes in effective tax rates. Therefore, the relation between firm size and the effective tax rate in the structural equations may not be contemporaneous. To address this potential lag problem, the averages of data over the sample period are used as crude measures.

---

[3] A reviewer expressed concern that NOLs may occur as a result of tax preferences, contrary to the hypothesis tested here. To check, a correlation was computed between NOLs and aggregate tax preference items (see fn. 8 below). The correlation was 0.047 and was not statistically significant.

A latent-variable approach is employed to mitigate the interpretation and measurement problems that arise from using different proxy measures. The latent-variable approach assumes that the variables in the structural equations are unobservable theoretical constructs (latent variables) that are measured by the empirically observable variables with error. A measurement model then specifies the relationship between each observable variable and its corresponding construct (e.g., Lambert and Larcker 1987). This approach addresses the measurement problems by explicitly considering the errors-in-variables problems and producing a single set of results for the relations between latent variables. The model measures *NOLS* by the frequency of net operating losses (*NOLFQY*) in the sample period.[4]

$$NOLFQY = \lambda_1 NOLS + \delta_1, \tag{3}$$

where λ_1 is factor loading and δ_1 is a measurement error.[5] The occurrence and utilization of NOLs were identified by examining tax footnotes to the financial statements.

ETR is measured by three different tax rates that prior studies have identified as income measures for the effective tax rate computation: economic income tax rate, or *ECORAT* (Siegfried 1972; USDT 1978), gross profit tax rate, or *CSHRAT* (Zimmerman 1983), and accounting income tax rate, or *ACCRAT* (Citizens for Tax Justice 1985; Joint Committee on Taxation [U.S. Congress] 1983; *Tax Notes* 1982).[6] Each of these measures is related to *ETR* as follows:

$$ECORAT = \lambda_2 ETR + \delta_2, \tag{4}$$

$$CSHRAT = \lambda_3 ETR + \delta_3, \tag{5}$$

$$ACCRAT = \lambda_4 ETR + \delta_4, \tag{6}$$

where λs are factor loadings and δs are measurement errors.[7]

ECORAT is the sum of total taxes paid (federal, state, and foreign) divided by the sum of "economic income" over the sample period. Economic income is measured as the sum of total tax preferences and taxable income. The current portion of the tax expenses reported in the financial statements is assumed to proxy the taxes paid. Pretax accounting earnings subtracting the increase in the deferred tax account "grossed up" by the statutory corporate tax rate is used to approximate taxable income.

Tax preferences are defined according to the Internal Revenue Code; that is, as the sum of certain exemptions, excess deductions, and income deferrals on which firms are subject to corporate minimum tax.[8] Because the items included in the list of tax preferences in the Internal Revenue Code are frequently revised, the tax preference list in the 1982 Tax of Equity and Fiscal Responsibility Act (TEFRA) is used.[9] In addition, long-term contracts and installment sales, which were not part of the 1982 TEFRA list, are included. Each tax preference item is measured by its corresponding timing or its permanent-difference item disclosed in the tax footnotes to the financial statements. For example, the depreciation tax preference is defined as the accelerated depreciation deductions for tax purposes over the straight-line depreciation, which is the method used by most firms in financial reporting. Thus, the depreciation tax preferences are roughly equal to the depreciation timing differences shown in the tax footnote to the financial statements.

---

[4] Another possible measure of *NOLS* is the amount of net operating losses. This measure is not used because the occurrence of net operating losses usually reduces tax rates to zero. More operating losses will not reduce tax rates further. Therefore, the relation between the amount of net operating losses and tax rates may not be meaningful.

[5] This is a typical way to set up a measurement equation in factor analysis with the unobservable variable on the right-hand side of the equation. In the estimation, λ_1 is constrained to 1 and δ_1 is set to 0 because, when the construct has only a single observed variable, it is assumed that there are no errors in measurement ($\delta_1 = 0$) and the factor loading is 1 ($\lambda_1 = 1$).

[6] Tax rates defined in Stickney and McGee (1982) were not examined in this study because the numerator of tax rates is roughly equal to taxable income multiplied by the corporate statutory tax rate minus tax credits. As a result, the only factor to account for the variations in tax rates computed as a ratio of taxes paid to taxable income in Stickney and McGee (1982) is the amount of tax credits. Therefore, tax rates computed by taxable income do not reflect the level of tax preferences and firms' success in the political process. Because of this limitation, this definition of tax rates is excluded from the analysis.

[7] There are three measurement equations because there are three measures for *ETR* with each measuring *ETR* with errors.

CSHRAT is the ratio of the sum of total taxes paid to the sum of gross profit (sales less cost of goods sold) over the sample period. *CSHRAT* is similar to the cash-flow tax rate in Zimmerman (1983). *ACCRAT* is the ratio of the sum of total taxes paid to the sum of pretax accounting earnings over the sample period. Pretax accounting income is the before-taxes book earnings reported by a firm in its financial statements. This effective tax rate is used in studies by *Tax Note* (1982), the Joint Committee on Taxation (1983), and Citizens for Tax Justice (1985).

Finally, the average total sales (*SALES*) and the average total assets (*ASSET*) during the sample period are used to measure *SIZE*.

$$SALES=\lambda_5 SIZE+\delta_5, \tag{7}$$

$$ASSET=\lambda_6 SIZE+\delta_6, \tag{8}$$

where λs are factor loadings and δs are measurement errors.

The above measurement model and the structural equation model (the path model) are jointly called the covariance structure model, which is estimated by the LISREL program developed by Jöreskog and Sörbom (1989). Figure 1 diagrams the overall covariance structure model. The latent variables are shown in circles and the proxies of the latent variables are shown in boxes.

Figure 1 posits that *SIZE* affects *ETR* both directly and indirectly through *NOLS*. If the overall relationship between *SIZE* and *ETR* is estimated alone without considering the path through *NOLS*, the estimate contains the effect of a variable that is irrelevant to firms' political success.

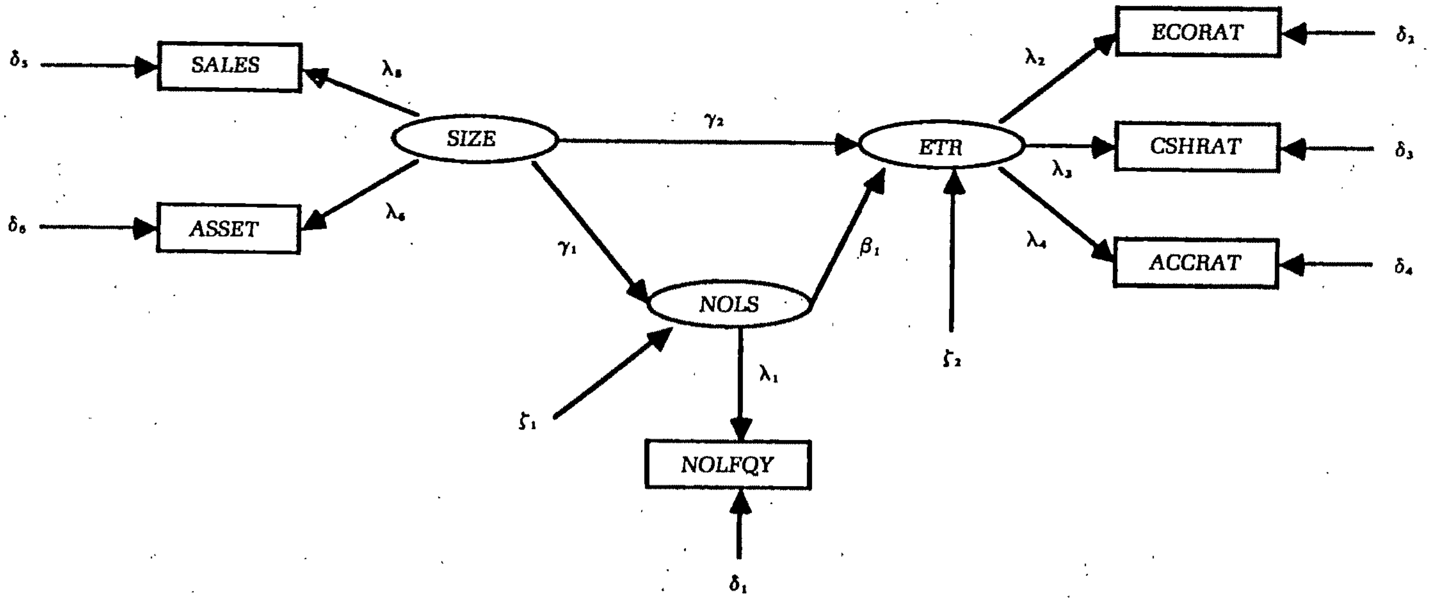
---

[8] Tax preferences include all special exemptions from taxable income, deductions in excess of normal amounts, and deferrals of taxable income that have enabled some individuals or corporations with substantial income to pay little or no tax while other taxpayers with less income have been required to pay a relative higher percentage of their income in taxes. However, Congress has traditionally singled out only those exemptions, excess deductions, and income deferrals deemed the most abused as tax preferences on which additional taxes may be imposed as defined in the Internal Revenue Code.

[9] The Internal Revenue Code Sections 57 and 291 define the items of tax preference relevant to corporate minimum tax. Items of tax preference in Section 57 include exclusion of interest and dividends; accelerated depreciation on real property; accelerated depreciation on leased personal property; amortization of certified pollution control facilities; mining exploration and development costs; circulation, research, and development costs; depletion; capital gains; incentive stock options; and intangible drilling cost. Section 291 includes interest deductions of financial institutions attributable to acquiring or carrying tax-exempt obligations and deemed dividends from international sales corporations.

Figure 1

The Path Model and the Measurement Model Estimating the Indirect Effect of NOLs on the Overall Firm Size and Effective Tax Rates Relation



Note: The latent variables (in circles) are *SIZE* (firm size), *NOLS* (net operating losses), and *ETR* (effective tax rate). Proxies for the latent variables (in boxes) are *SALES* (average total sales), *ASSET* (average total assets), *NOLFQY* (frequency of net operating losses), *ECORAT* (economic income tax rate, *CSHRAT* (gross profit tax rate), and *ACCRAT* (accounting income tax rate). The parameters are β_1, γs (regression coefficients), λs (factor loadings), δs (measurement errors), and ζs (errors in equations).

Table 1

Industries Selected, Sample Size, and the Size Distribution of the Sample Firms

Industry	*SIC #*	*Maximum Sales (in $million)*	*Minimum Sales (in $million)*	*Average Sales (in $million)*	*Sample Size*
Petroleum refining	291	88,561.0	5.3	8,291.2	35
Aerospace and defense	372, 378	14,669.0	4.9	1,326.9	34
Office computing/accounting equipment	357	40,180.0	3.2	1,648.2	43
Retailing	531, 533	35,883.0	35.2	2,380.4	31

Note: The SIC number is taken from the NAARS database.

Sample Selection

To ensure that firms with both high and low political influence and costs were examined in the analysis, the sample included firms from two industries that are considered to be highly sensitive politically and two other industries that are considered to have low political sensitivity. The petroleum refining (oil) industry (SIC 291) and the aerospace and defense industry (SIC 372 and 378) were included to represent firms with high political costs, while the retailing industry (SIC 531 and 533) and the office computing/accounting equipment industry (SIC 357) were included to represent firms with low political costs. The NAARS database was used to identify all the firms in these selected industries, then sample firms were randomly chosen for each industry. Data for the six-year period 1978–1983 were collected for the sample firms. Table 1 summarizes the industries selected, number of firms, and the average size distribution in each industry for the six-year period. Table 2 contains descriptive statistics of the variables in the measurement model.

II. Data Analysis

Correlation Analysis

Table 3 presents the product moment correlations for the observed variables. *NOLFQY* is shown to have had high negative correlations with all three tax rates. This indicates that the occurrence of NOLs has a substantial effect on effective tax rates and raises the omitted-variables problem for previous research that estimated only the overall firm size and tax rates relation.

Covariance Structure Analysis

Because *NOLFQY* variable is ordinal, which violates the multivariate normality assumptions by the maximum likelihood estimate, the approach suggested by Jöreskog and Sörbom (1985) in LISREL regarding the analysis of ordinal data was followed.[10]

---

[10] This approach involves estimation of the polyserial correlations and analysis of the matrix of such correlations using the unweighted least squares method (ULS). By assuming that an ordinal variable is a crude measure of an underlying, unobserved continuous variable that is normally distributed with zero mean and unit variance, the polyserial correlation is computed as the correlation between the underlying unobserved continuous variable and another continuous variable measured on an interval scale (Jöreskot and Sörbom

Table 2
Descriptive Statistics of the Observed Variables in the Measurement Model

	Mean	*Standard Deviation*
ECORAT	0.248	0.188
CSHRAT	0.080	0.086
ACCRAT	0.241	0.199
SALES ($ million)	3,710.0	10,670.0
ASSET ($ million)	2,680.5	7,299.0

Frequency of NOLs (NOLFQY)	
0	56
1	12
2	18
3	14
4	15
5	7
6	21

Note: *ECORAT*=economic income tax rate; *CSHRAT*=gross profit tax rate; *ACCRAT*=accounting income tax rate; *SALES*=total sales; *ASSET*=total assets.

The polyserial correlation matrix used in the analysis is reported in table 3.[11] Table 4 presents results of the analysis of the polyserial correlation matrix by the ULS method.[12] It contains the goodness-of-fit indices for the overall model and related statistics for the measurement model. Similar statistics for the analysis of the product moment correlation matrix (by the ULS method) are also included in table 4. Only results of the former approach will be discussed.

The χ^2 value for ULS is 7.28 ($p=0.507$), indicating that the hypothesized model can generate the observed covariance matrix and the model is a good fit. It is a test against the alternative hypothesis that the fitted covariance structure is unconstrained. The observed probability level associated with the χ^2 value is greater than the standard cutoff of 0.01 or 0.05. Therefore, the hypothesis that the theoretical model can explain the covariance structure of the observed variables cannot be rejected.

1988). When two variables are measured on an interval scale, the polyserial correlation between them is identical to the product moment correlation.

Although ULS allows more general assumptions regarding the distribution of variables, it should be noted that the chi-square, standard errors, and t-values estimated by ULS are still computed under the multivariate normality assumptions. Therefore, while the parameter estimates of ULS are consistent, the parameter estimates and related statistics may not be asymptotically efficient.

[11] The polyserial correlations are estimated by the PRELIS program developed by Jöreskog and Sörbom (1988).

[12] The use of the correlation matrix in the analysis assumes that the latent variables have zero mean and unit variance. This assumption is made because the scales of the variables are not essential to the tests proposed in this study.

Table 3

Product Moment Correlation Matrix and Polyserial Correlation Matrix for the Observed Variables in the Measurement Model

	NOLFQY	ECORAT	CSHRAT	ACCRAT	SALES
Product Moment Correlations					
ECORAT	−0.605				
CSHRAT	−0.522	0.714			
ACCRAT	−0.603	0.953	0.713		
SALES	−0.181	0.165	0.304	0.161	
ASSET	−0.191	0.155	0.295	0.149	0.971
Polyserial Correlations					
ECORAT	−0.619				
CSHRAT	−0.563	0.714			
ACCRAT	−0.612	0.953	0.713		
SALES	−0.216	0.165	0.304	0.161	
ASSET	−0.218	0.155	0.295	0.149	0.971

Note: *ECORAT*=economic income tax rate; *CSHRAT*=gross profit tax rate; *ACCRAT*=accounting income tax rate; *SALES*=total sales; *ASSET*=total assets; *NOLFQY*=frequency of NOLs.

The goodness-of-fit index (GFI) measures the relative amount of variances and covariances explained by the model and takes on a value between 0 and 1. The GFI value of 0.996 reported in table 4 supports the conclusions based on the χ^2 test.

The factor loading of each observed variable on its corresponding construct (λ) and its associated z-statistic are also reported in table 4.[13] Each λ is positive and close to one and the associated z-statistic is large, indicating that each observed variable has a positive and statistically significant relation with its corresponding proposed construct. The squared multiple correlation provides a measure of the strength of the linear relationship for each structural equation, and the coefficient of determination is a measure of the strength of the relationships for all structural equations jointly. The squared multiple correlation value is 0.048 for equation (1) and 0.518 for equation (2). The coefficient of determination for all structural equations is 0.066. This suggests that the explanatory power of each structural equation and the overall structural model is nontrivial.

The parameter estimates for equations (1) and (2) are reported in table 5. With the exception of the *SIZE* and *ETR* relation (γ_2), each coefficient in the structural equations is statistically significant at the 0.05 level. The negative relation between *SIZE* and *NOLs* indicates that small firms are more likely to incur net operating losses. *NOLs* and *ETR* also are inversely related, which indicates that the occurrence of net operating losses significantly reduces *ETR*. The tendency of small firms to incur net operating

---

[13] λ_1 is constrained to 1 because only one proxy, *NOLFQY*, is proposed for the *NOLs* construct. λ_5 of ULS is also constrained to 1 because the unconstrained estimation produces an "improper solution" (i.e., a standardized loading greater than 1 and a negative measurement error variance). The procedure suggested by Gerbing and Anderson (1987) is followed by constraining λ_5 to be 1 and reestimating the respecified model. No z-statistics are reported for those λs constrained to 1.

Table 4

Parameter Estimates of the Measurement Model and Goodness-of-Fit Indexes for the Overall Covariance Structure Model

	ULS-Polyserial Correlations[a]	*ULS-Product Moment Correlations*[b]
Goodness-of-Fit Indexes		
Chi-square (χ^2)	7.28	8.76
Probability level	0.507	0.363
GFI index	0.996	0.996
Square Multiple Correlations		
Equation (1)	0.048	0.036
Equation (2)	0.518	0.484
Coefficient of Determination	0.066	0.062
Measurement Model		
NOLS *construct*		
NOLFQY (λ_1)	1.000[c]	1.000
	—	—
ETR *construct*		
ECORAT (λ_2)	0.823	0.836
	(27.45)[d]	(27.17)
CSHRAT (λ_3)	0.871	0.856
	(8.38)	(7.91)
ACCRAT (λ_4)	0.816	0.832
	(26.96)	(27.45)
SIZE *construct*		
SALES (λ_5)	1.000	1.000
	—	—
ASSET (λ_6)	0.968	0.969
	(8.90)	(8.85)

[a] Polyserial correlation matrix estimated by ULS.

[b] Product moment correlation matrix estimated by ULS.

[c] λ was constrained to one.

[d] z-statistics computed as a ratio of standardized factor loadings to standard errors.

losses, which in turn reduces tax rates, suggests that *NOLs* tend to cause small firms to have lower tax rates. Consequently, this effect of *NOLs*, if not controlled, will cause a positive *SIZE* and *ETR* relation. The extent of such an effect of *NOLs* is evaluated by path analysis.

The model depicted in figure 1 has two paths. The first is an indirect path from *SIZE* through *NOLs* to *ETR*, which represents the part of the overall *SIZE* and *ETR* relation that is attributable to *NOLs* (i.e., the indirect effect of *NOLs*). The second is a direct path from *SIZE* to *ETR* after controlling the indirect effect of *NOLs*. In other words, path analysis decomposes the total effect of *SIZE* on *ETR* into that attributable to *NOLs* and that unrelated to *NOLs*.

The indirect effect of *NOLs* (0.125) is a segment of total effect of *SIZE* on *ETR* attributable to NOLs and is equal to the product of the coefficient of the *SIZE* and *NOLs* relation (−0.220) and the coefficient of the *NOLs* and *ETR* relation (−0.569). The indirect effect of *NOLs* is statistically significant at the 0.05 level. This positive indirect effect of

Table 5
Parameter Estimates of the Structural Equations and Direct and Indirect Effects Analysis
(t-statistics in parentheses)

	ULS-Polyserial Correlations[a]	*ULS-Product Moment Correlations*[b]
Parameter Estimates		
SIZE→NOLs (γ_1)	−0.220	−0.188
	(−2.68)*	(−2.30)*
NOLs→ETR (β_1)	−0.569	−0.653
	(−5.75)*	(−6.13)*
SIZE→ETR (γ_2)	0.081	0.103
	(1.27)	(1.49)
Effect Analysis		
Indirect Effect of NOLs[c]	0.125	0.104
	(2.40)*	(2.21)*
Direct Effect[d]	0.081	0.103
	(1.27)	(1.49)
Total Effect[e]	0.206	0.207
	(2.75)*	(2.69)*

* Statistically significant at the 0.05 level.

[a] Polyserial correlation matrix analyzed by ULS.

[b] Product moment correlation matrix analyzed by ULS.

[c] The extent of the total effect of *SIZE* on *ETR* that is attributable to *NOLs*.

[d] The direct effect of *SIZE* on *ETR* after the indirect effect of *NOLs* is removed. The direct effect is equal to the *SIZE* and *ETR* relation in the structural equation (γ_2).

[e] The total effect of *SIZE* on *ETR* without removing the indirect effect of *NOLs*. The total effect should be equal to the sum of the indirect effect of *NOLs* and the direct effect.

NOLs is consistent with previous analyses of the individual parameters of the structural equations showing that the tendency of small firms to incur net operating losses causes a positive *SIZE* and *ETR* relation.

The total effect of *SIZE* on *ETR* (0.206), which is statistically significant at the 0.05 level, consists of 0.125 indirect and 0.081 direct effects. This indicates that if the total effect of *SIZE* on *ETR* is used to assess the firm size and political success relation, as was done in prior studies, the relation may be overstated, and inappropriate conclusions may be drawn.

III. Conclusions

A path model with a latent-variable approach was developed and estimated to show the indirect effect of NOLs on the overall relationship between firm size and effective tax rates. The results indicate that a significant portion of the overall firm size/tax rate relation was attributable to the indirect effect of NOLs. Since NOLs result mainly from firms' operating characteristics rather than firms' political environment, the use of total effect of firm size on tax rates in previous studies to test the association of firm size and political success may be biased and can lead to improper conclusions.

Certain limitations may apply to this study. Due to the sample selection process employed, some results of the analysis may be sensitive to sample variations. For example, the firm size relations with NOLs and effective tax rates may be sensitive to the characteristics of the firms included in the sample. In addition, although the unweighted least square method (ULS) used in the analysis allows more general assumptions regarding the distribution of variables, the standard errors estimated under the ULS are still computed under multivariate normality assumptions. In the absence of multivariate normality for the data (the ordinal variable of *NOLFQY*), the standard errors may be understated. This may, in turn, overstate the statistical significance of the tests.

References

Citizens for Tax Justice. 1985. Corporate Taxpayers and Corporate Free Loaders: Four Years of Continuing Legalized Tax Avoidance by America's Largest Corporations. Washington, D.C.

Duncan, O. D. 1975. *Introduction to Structural Equation Models.* New York: Academic Press.

Gerbing, D. W., and J. C. Anderson. 1987. Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternative respecifications. *Psychometrika* 52 (March): 99–111.

Jöreskog, K. G., and D. Sörbom. 1985. *LISREL VI: Analysis of Linear Structure Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods.* Mooresville, IN: Scientific Software, Inc.

———, and ———. 1988. *PRELIS: A Program for Multivariate Data Screening and Data Summarization.* 2nd ed. Mooresville, IN: Scientific Software, Inc.

———, and ———. 1989. *LISREL 7: A Guide to the Program and Applications.* 2nd ed. Chicago: SPSS, Inc.

Kmenta, J. 1971. *Elements of Econometrics.* New York: Macmillan.

Lambert, R. A., and D. F. Larcker. 1987. An analysis of the use of accounting and market measures of performance in executive compensation contracts. *Journal of Accounting Research* 25 (Supplement): 85–125.

Siegfried, J. J. 1972. The relationship between economic structure and the effect of political influence: Empirical evidence from the corporation income tax program. Ph.D. dissertation, University of Wisconsin.

Stickney, C. P., and V. E. McGee. 1982. Effective corporate tax rates: The effect of size, capital intensity, leverage, and other factors. *Journal of Accounting and Public Policy* 1 (Winter): 125–52.

Tax Notes. 1982. Effective corporate tax rates in 1980. *Tax Notes* 14 (March): 889–944.

U.S. Congress, Joint Committee on Taxation. 1983. Study of 1982 effective tax rates of selected large U.S. corporations. Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Treasury. 1978. Effective income tax rates paid by United States corporations in 1972. Washington, D.C.: U.S. Government Printing Office.

Zimmerman, J. L. 1983. Taxes and firm size. *Journal of Accounting & Economics* 5 (August): 119–49.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 170–187

Tradeoffs in the Choice Between Logit and OLS for Accounting Choice Studies

Mary Stone
University of Alabama
John Rasp
Stetson University

SYNOPSIS: Many accounting studies examine dichotomous choices (e.g., qualify/do not qualify an audit opinion or capitalize/do not capitalize a cost). These studies often involve small overall sample sizes, disparate response group sizes, and predictor variables that are skewed and collinear. These factors can cause distributional problems in test statistics for logit (or probit) regression models, which can lead to incorrect inferences. However, few empirical benchmarks exist for assessing the effect of these factors.

The current paper determines how response group size and the number, distribution, and correlation of predictor variables affect empirical error rates and the minimum required sample size for using logit. Comparisons are made with the error rates obtained from an ordinary least squares (OLS) linear probability model, an alternative that has been suggested for small sample studies. Because accounting researchers choosing between logit and OLS may be concerned with more than the calibration of the models' test statistics, comparisons of the sensitivity of logit and OLS parameter estimates to the range of data sampled for the predictor variables and of the models' classificatory ability also are made.

Both simulated and real accounting data are used. The results of Monte Carlo simulations show that logit test statistics are biased when the sample size is small. However, much of the bias is attributable to the skewness of the predictor variables, a problem that is characteristic of accounting research and that also affects OLS test statistics. In such set-

Support for this project was provided by Peat, Marwick, Main, by a College of Commerce Summer Research Fellowship, and by the Center for Current Accounting Issues at the University of Alabama. The helpful suggestions of Mike Dugan, Rob Ingram, Mike Roberts, and anonymous referees are acknowledged.

Submitted August 1988.
Revisions received May 1989, February and July 1990.
Accepted September 1990.

tings, OLS may result in test statistics that are minimally better calibrated. The parameter estimates of the model, however, will be more sensitive to the sampling frame. Furthermore, experimentation with data on auditors' Statement No. 87 consistency judgments indicates that OLS also may result in higher Type I error rates when it is used for prediction or classification. These results are interpreted as indicating that, even for sample sizes as small as 50, logit rather than OLS still may be the preferable model for accounting choice studies.

Key Words: *Empirical error rates, Sample size, Response group size, Multicollinearity.*

Data Availability: *The data and simulation programs used in this study are available from the authors.*

MANY of the decisions of interest to accounting researchers involve explanation or prediction of mutually exclusive, dichotomous choices (e.g., qualify/do not qualify an audit opinion, capitalize/do not capitalize a cost). Most recent papers examining such choices use dichotomous logistic (logit) or dichotomous probit regression rather than ordinary least squares (OLS) regression.[1] Although these papers seldom provide a rationale for the estimation method used, the earliest papers using logit or probit (e.g., Hagerman and Zmijewski 1979) cite the inefficiency of OLS parameter estimates for models with dichotomous response variables. Additional theoretical and econometric reasons for preferring logit or probit to OLS when the response variable is dichotomous and the sample size is large are provided in the statistics literature (e.g., Amemiya 1981; McFadden 1982).

However, the choice of model is not as clear when the sample size is small. The statistics literature provides little guidance. A recent study of the effect of overall sample size on empirical error rates for probit (Noreen 1988) concludes that for sample sizes of fewer than 100, probit offers no real advantage over OLS.

In this paper, we examine the effects of sample size, number, correlation, and distribution of predictor variables on logit error rates.[2] Comparisons of the sensitivity of logit and OLS parameter estimates and of the models' ability to predict auditors' consistency judgments also are made. The objective is to provide empirical benchmarks that accounting researchers can refer to in assessing the extent to which commonly encountered problems are likely to affect the inferences that can be drawn from logit or OLS test statistics.

[1] This statement was based on our review of articles published in *The Accounting Review*, the *Journal of Accounting Research*, the *Journal of Accounting and Public Policy*, and the *Journal of Accounting & Economics* from 1986–1988. Ten of the articles used logit and nine used probit to study dichotomous choices. None used only OLS. One (Shevlin 1987) noted model estimation using OLS as well as logit but only reported logit results. Another (Whittred 1987) reported results for OLS and probit. None of the studies reported results for both logit and probit. The studies are listed in appendix A.

[2] Logit uses the cumulative logistic probability function to transform the predictor variables, while probit uses the normal probability function. Maddala (1983) and Amemiya (1981) demonstrate the comparability of the results of dichotomous logit and probit models, with the latter noting that this comparability does not always hold for multichotomous models. The use of logit in the current study should be viewed as a reflection of software availability rather than as an important element of the research design.

The remainder of the paper is organized as follows. In the next section, we (1) briefly review the differences between logit and OLS, (2) identify the data problems that affect logit sample size requirements, and (3) establish the prevalence of these problems in accounting research. In section II, we investigate the sample size requirements of logit and OLS test statistics. In section III, we discuss other factors relevant to the choice between logit and OLS—namely, the effect of different sampling frames, classificatory accuracy, and the power of the tests. The final section provides a summary and conclusions.

I. Differences Between Logit and OLS

Consider a researcher seeking to predict a dichotomous ("failure/success") outcome variable, Y_i, as a function of one or more predictor variables X_{ij}.

Use of standard regression techniques results in modeling the response as a linear function of the predictor variables, Thus,

$$Y_i = \sum_j \beta_j^* X_{ij} + \epsilon_i, \tag{1}$$

where the β_j^*s are parameters describing the relationship between X and Y and the ϵ_is are error terms, distributed independently with 0 mean. Standard regression procedures produce ordinary least squares (OLS) estimates of the parameters, $\hat{\beta}_j^*$. The probability, P_i, of observing a "success" on subject i, given predictors X_{ij}, can then be estimated by:

$$P_{i,\mathrm{OLS}} = \sum_j \hat{\beta}_j^* X_{ij}. \tag{2}$$

The logit procedure models the probability of "success" by:

$$\ell n(P_i/(1-P_i)) = \sum_j \beta_j X_{ij}. \tag{3}$$

The parameters generally are estimated by maximum likelihood techniques. Using the logit approach, the probability of observing a "success" on subject i may be estimated by:

$$P_{i,\mathrm{Logit}} = \exp\left\{\sum_j \hat{\beta}_j X_{ij}\right\} \bigg/ \left(1 + \exp\left\{\sum_j \hat{\beta}_j X_{ij}\right\}\right). \tag{4}$$

The logit and OLS procedures differ in their underlying assumptions. The OLS test statistics require that the error terms (ϵ) be normally distributed with constant variance. Because they are not when the dependent variable is dichotomous, the use of OLS results in inefficient parameter estimates.[3]

Because OLS uses a linear function to model possibly nonlinear relationships, it can produce meaningless predicted probabilities of greater than one or less than *zero.* The sensitivity of its parameter estimates to the range of data sampled has led to recommendations that probit or logit rather than OLS be used (Hanushek and Jackson 1977, 210).

---

[3] This point is demonstrated in Aldrich and Nelson (1984, 13).

Logit uses a (nonlinear) cumulative logistic probability function to transform the predictor variables. The functional form of the model facilitates analysis of the effects of changes in the values of the predictor variables on the probability estimates that is not possible with OLS.[4] It also automatically produces probability estimates that fall between 0 and 1, and that Collins and Green (1982) assert result in a smaller Type I error rate than does OLS.[5]

Logit has the favorable features noted above for any sample size. However, the statistical theory that establishes desirable properties for maximum likelihood estimates does so only asymptotically (Aldrich and Nelson 1984, 53). Little guidance is given on how large the sample must be for logit parameter estimates to be unbiased or for tests based on asymptotic normal theory to be valid. The issue is complicated by the interactive effect of response group size and the number, distribution, and correlation of predictor variables.[6]

Aldrich and Nelson (1984, 81) conclude that the choice between logit and probit or OLS for small samples "depends upon the number of variables, the degree of covariation, and so forth, but for most applications, the rule of thumb of 50 cases per parameter seem(s) safe." Pindyck and Rubinfeld (1981, 290) note the relevance of response group, as well as overall sample size.

A review of 19 accounting studies using logit or probit (see appendix A) indicates that small sample sizes, disparate response group sizes, and skewed and collinear predictor variables are prevalent in accounting research. Samples ranging from 29 to 800 observations were used. When total sample size is divided by the number of parameters estimated, only four articles (Dopuch et al. 1987; Palepu 1986; Francis 1987; El-Gazzar et al. 1986) had at least one model with the 50 total observations per parameter identified as a rule of thumb for judging the adequacy of sample size (Aldrich and Nelson 1984). Nine of the 19 studies estimated models using less than 100 total observations.

As the last column of the appendix indicates, response group sizes for the studies seldom were equal. The number of observations from the smaller group per predictor often was small. Thirteen of the studies had at least one model with ten or fewer observations from the smaller group per predictor. Of these, four had three or fewer observations from the smaller group per predictor.

The predictor variables in these studies frequently included firm size, leverage, profitability, and related ratios. The distributions of these and other ratios tend to be skewed (Frecka and Hopwood 1983; Hopwood et al. 1988). The bivariate correlations for predictor variables in the appendix studies varied from 0.09 to 0.63 (e.g., Casey et al. 1986).

The foregoing review identifies small sample size, disparate response groups, and skewed and collinear predictor variables as problems confronting accounting researchers. To date, only one of these issues has been examined using accounting data.

---

[4] For the OLS model, we have $p=\beta_0+\beta_1X_1+\ldots+\beta_kX_k$, and hence $\partial p/\partial X_1=\beta_1$, which is constant for all values of X_1. For the logit model, $p=E/(1+E)$, where $E=\exp\{\beta_0+\beta_1X_1+\ldots+\beta_kX_k\}$. Thus $\partial p/\partial X_1=\beta_1E/(1+E)^2$, which varies depending not only on the level of X_1, but also on the values of the other predictor variables.

[5] Collins and Green's (1982, 352) assertion is based on experimentation with a hypothetical data set that they argue was representative of bankruptcy prediction and other problems in which "one group's value is typically clustered around a certain value while the other group has a much larger variance."

[6] Discussion of the effect of these factors on logit and probit sample size requirements can be found in Aldrich and Nelson (1984), Hanushek and Jackson (1977), Maddala (1983), and Pindyck and Rubinfeld (1981).

Noreen (1988) examines the effect of overall sample size on empirical error rates for probit and OLS. His results indicate that for sample sizes of 100 or fewer (1) the empirical error rate for the test of the overall significance of the probit model is approximately twice its nominal rate, and (2) the empirical error rate for the t-test for the individual parameters of the model tends to be significantly lower than the nominal rate. The remainder of this paper examines the effect of overall sample size on logit rather than probit error rates. It also assesses the effect of unequal response group sizes and skewed and collinear predictor variables, problems not examined by Noreen (1988).

II. Logit Sample Size Requirements

The research described in this section provides empirical benchmarks for assessing logit sample size requirements.

Test Statistics and Empirical Error Rates

Simulation is used to assess the effect of (1) overall sample size, (2) disparity of response group sizes, and (3) the number, skewness, and distribution of predictor variables. The following test statistics are examined.

1. Values of t-statistics testing individual model parameters, where H_0: $\beta=0$ and the test statistic, $t=\hat{\beta}/\text{s.e.}\ (\hat{\beta})$, is compared with a student's t distribution with $n-k-1$ degrees of freedom; this is done for both logit and OLS models.
2. Values of the chi-square statistic for the overall logit model, where H_0: $\beta_1=\beta_2=\ldots=\beta_k=0$ and a likelihood ratio test is used. The test statistic is compared with the chi-square distribution with k degrees of freedom.
3. Values of the F-statistic for the overall OLS model, where H_0: $\beta_1=\beta_2=\ldots=\beta_k=0$. The test statistic is compared with the F distribution with k and $n-k-1$ degrees of freedom.

The use of simulation allows us to experiment efficiently with a number of alternative scenarios using 10,000 replications and to isolate the effect of individual data problems on sample size requirements.[7] The steps taken in the simulation and the analysis of results are as follows.

Step 1. "X" observations for k ($k=1$ and 3) continuous predictor variables and one indicator (dummy) variable for samples of 50 and 100 are generated. Both normal and right-skewed distributions of the predictors are considered. Also considered are data with no correlation ($\rho=0$) and with moderate correlation ($\rho=0.5$) among the k variables. The correlation between the continuous variables and indicator variable is zero throughout the analysis. The indicator variable is randomly generated by assigning a value of 1 with probability 0.6 and a value of 0 with probability 0.4. Skewness is introduced by generating normal data using the GGNSM subroutine of the IMSL mathematical and statistical software package and then exponentiating to get lognormal data. Lognormal data with correlation $\rho=0.50$ are obtained by generating and exponentiating normal data with correlation $\rho^*=\ell n\{\rho(e-1)+1\}$.[8]

---

[7] The simulations were programmed in Fortran 77 using standard computational procedures. Computational accuracy was checked by comparison with results from the SAS logistic and regression procedures. Because of the number of replications, it was more efficient to use our Fortran programs than to run SAS repeatedly.

[8] We also examined sample sizes of 30 and 200, six and nine predictors, and $\rho=0.90$. The results for these additional experiments are consistent with those reported.

Step 2. "Y" 0/1 response variables are generated, for population success probabilities of $p=0.5$, 0.7, and 0.9.[9] Because the behavior of the test statistic is being investigated under the null hypothesis of no relation between response and predictor variables, data for "Y" and "X" are generated independently.

Step 3. The "X" predictors generated in step 1 are used to fit the "Y" responses generated in step 2. The relation is modeled using both logit and OLS.

Step 4. The frequency with which the null hypothesis would be rejected is determined. This is done using each of the estimated test statistics (t, χ^2, and F) for different commonly selected nominal significance (alpha) levels (0.1, 0.05, and 0.01) and for the different sample sizes, response group sizes, and distributions and correlations of predictor variables.

Simulation Results

Table 1 is based on 10,000 replications of a model with one normally distributed predictor variable. The null hypothesis of no significant effect is true for the parameter estimate and for the overall model. The first column of the table shows the nominal significance level for testing the null hypothesis of no effect of the predictor variable. The column labeled t-test shows the frequency with which the null hypothesis actually was falsely rejected when the parameter was estimated by logit and by OLS (i.e., the empirical error rate). The final column reports the false rejection rate for the test of the overall logit model (i.e., chi-square test) and for the F-test of the OLS model.

The first row of entries shows that the logit t-test is conservatively biased, falsely rejecting the null hypothesis less often than indicated by the nominal significance level. In contrast, the test of the overall logit model is anticonservatively biased. The difference between empirical and nominal error rates for both tests is smaller for higher nominal significance levels. At $n=50$, for example, the difference between the 1 percent nominal and 0.34 percent empirical error rates for the logit t-test is 66 percent of the nominal significance level. For the same sample size and a nominal significance level of 10 percent the difference is only 11.2 percent of the nominal significance level. The pattern is similar for the test of the overall logit model, decreasing from a difference that is 23 percent of the 1 percent nominal significance to a difference that is 10.3 percent of the 10 percent significance level.

Increasing the sample size also decreases the difference between nominal and empirical error rates. For example, for the t-test at $n=50$ the difference between the 5 percent nominal and 3.67 percent empirical error rate is 1.33 (26.6 percent). For the same nominal level but a sample size of 100 the difference decreases to 0.74 (14.8 percent). Results not reported indicate that the difference for the logit t-test are not statistically significant for sample sizes of 200 or more.

A similar comparison for the test of the overall logit model at a 5 percent nominal significance level indicates a decrease from a difference of 0.64 (12.8 percent) at $n=50$ to a difference of 0.19 (3.8 percent) at $n=100$. (Notably, differences between nominal and empirical significance levels for this test are not statistically significant at $n=100$.

In contrast to some of the logit results, the differences between the nominal and empirical error rates for OLS are no larger than would have been expected because of sampling variation. In no case is the difference statistically significant.

[9] Because of symmetry, it is unnecessary to consider 0.1 and 0.3

Table 1

The Effect of Total Sample Size on Empirical Significance Levels for Logit and OLS Models with One Normally Distributed Predictor and Equal Response Group Sizes[a]

Sample Size = 50

Rejection Level	*Model*	*t-Test*	*Test of Overall Model*[b,c]
1%	Logit	0.34%*[d]	1.23%*
	OLS	0.87%	0.87
5%	Logit	3.67%*	5.64%*
	OLS	5.06%	5.06
10%	Logit	8.88%*	11.03%*
	OLS	10.38%	10.38

Sample Size = 100

Rejection Level	*Model*	*t-Test*	*Test of Overall Model*
1%	Logit	0.60%*	1.06%
	OLS	0.88	0.88
5%	Logit	4.26%*	5.19%
	OLS	5.09	5.09
10%	Logit	9.23%*	10.27%
	OLS	9.88	9.88

[a] Based on 10,000 replications.

[b] The test for the significance of the logit model was a chi-square test. The test for the significance of the OLS model was an F-test.

[c] With one predictor variable, the OLS t-test for an individual coefficient is equivalent to the OLS F-test of the overall model.

[d] The standard two-sample test for differences in proportions was used.

An * indicates cases in which the empirical significance level differs significantly from the nominal level (at $\alpha = 0.10$).

Table 1 presents the simplest case, to illustrate the fundamental behavior of the test statistics. Table 2 more closely approximates a typical accounting choice study. The simulations used three continuous predictor variables and one indicator variable. The panel A results for the continuous and indicator variables show that the logit t-tests are significantly conservatively biased at the 1 and 5 percent nominal significance levels but not at the 10 percent level. The difference between nominal and empirical significance levels always is greater for the continuous variables than for the indicator variable. The test of the overall logit model is anticonservatively biased at each of the nominal significance levels. In contrast, the OLS error rate differs from the respective nominal rate only for the F-test at the 5 percent significance level.

Comparing the t-tests in panel A of table 2 with those in table 1 shows that increasing the number of predictor variables generally results in an empirical error rate that is closer to the nominal error rate. At a sample size of 50 and a nominal significance level

Table 2

Comparison of Logit and OLS Empirical Error Rates Model with Three Continuous and One Categorical Predictor Variables, with Equal Response Group Sizes[a]

Panel A. Sample Size=50
Normally Distributed Predictors:

Rejection Level	Model	Tests of Individual Coefficients: Continuous	Tests of Individual Coefficients: Indicator[b]	Tests of Overall Model
1%	Logit	0.28%*[c,d]	0.54%*	1.23%*
	OLS	0.86	1.06%	0.84%
5%	Logit	4.06*	4.53*	6.21*
	OLS	4.92	4.91	4.60*
10%	Logit	9.70	10.18	12.10*
	OLS	10.02	9.90	9.67

Panel B. Sample Size=50
Skewed Continuous Predictors:

Rejection Level	Model	Tests of Individual Coefficients: Continuous	Tests of Individual Coefficients: Indicator	Tests of Overall Model
1%	Logit	0.07%*	0.58%*	1.79%*
	OLS	0.06*	1.02	0.71*
5%	Logit	2.13*	4.74	7.38*
	OLS	3.94*	5.18	3.60*
10%	Logit	6.93*	10.26	14.26*
	OLS	9.27*	10.20	8.43*

[a] Based on 10,000 replications.

[b] Coded one 60 percent of the time, zero otherwise.

[c] A standard two-sample test for differences in proportions was used. An * indicates cases in which the empirical significance level differs significantly from the nominal level (at $\alpha=0.10$).

[d] There are three continuous variables in each of the 10,000 replications. The empirical error rate is the percentage of these 30,000 tests that falsely rejected the null hypothesis.

of 5 percent, for example, the empirical error rate for the continuous variables increases from 3.67 to 4.06 percent. Results not shown document that as the number of continuous predictors increases from six to nine the empirical error rate increases from 4.60 to 5.20 percent, neither of which is significantly different from the 5 percent nominal significance level. This increase occurs because the model becomes more overfitted as additional predictors are added.

Comparing the results for the test of the overall logit model in panel A of table 2 with those in table 1 shows that increasing the number of predictors increases the difference between the nominal and empirical error rates. At a sample size of 50 and a nominal significance level of 5 percent, for example, the difference between the nominal and empirical error rates increases from 0.64 (12.8 percent of the nominal significance level) to 1.21 (24.2 percent of the nominal significance level).

Table 3
The Effect of Distribution and Relative Response Group Size on the Empirical Error Rate for Tests of the Overall Model[a] (Nominal Significance Level = 0.05)

		Normal Predictors K=3		*Skewed Predictors* K=3	
Sample Size	*Percent of Obs. in Group One*	*Logit*	*OLS*	*Logit*	*OLS*
n=50	50	6.45*[b]	5.15	7.58*	3.94*
	70	6.39*	5.45*	8.64*	3.91*
	90	7.39*	4.77	**[c]	8.09*
n=100	50	5.79*	5.24	6.70*	3.90*
	70	5.60*	5.41*	7.19*	4.63*
	90	5.93*	5.05	7.23*	7.29*

[a] Based on 10,000 replications.

[b] A standard two-sample test for differences in proportions was used. An * indicates cases in which the empirical significance level differs significantly from the nominal level (at $\alpha = 0.10$).

[c] **Maximum likelihood estimates frequently failed to converge.

Comparing the logit results in panels A and B of table 2 indicates that the difference between nominal significance levels and empirical error rates for tests of the logit model is greater when data are skewed. For example, when data are normally distributed (panel A), the difference between nominal and empirical error rates for the *t*-test on the continuous variables at a 5 percent nominal significance level is 0.94 (18.8 percent of the nominal level). When data are skewed (panel B), the difference increases to 2.87 (57.4 percent). Skewness also exacerbates the anticonservative bias of the test of the overall logit model. It increases differences between empirical and nominal error rates from 0.23, 1.21, and 2.1 at the 1, 5, and 10 percent significance levels for normally distributed data to 0.79, 2.38, and 4.26 for skewed data. For this test, as well as for the *t*-tests on the continuous variables, comparing nominal and empirical significance levels for skewed and normal data results in larger differences than do comparisons made when only sample size differs.

Notably, skewness also results in OLS empirical error rates that are significantly different from the respective nominal rates. The OLS *t*-tests for the continuous variables as well as for the overall F-test are conservatively biased. This result suggests that regardless of the model used, skewed predictor variables need to be identified and when possible transformed.[10]

Table 3 demonstrates the effect of unequal response group sizes on tests of the overall model.[11] Results are reported for cases where there are (1) an equal number of observations in the treatment and control groups (i.e., an equal number coded 0 and 1); (2) 70 percent of the observations are in one group and 30 percent in the other; and (3) 90 percent are in one group and 10 percent in the other. The data were uncorrelated. Normally distributed predictors were used to generate the results on the left side of the table while skewed data were used to generate those on the right.

---

[10] For a discussion of transformation corrections for nonnormality, see section 1.2 of Miller (1986). In using these corrections, it should be kept in mind that transformation of one variable will change its covariances with other variables.

[11] Results not shown indicate *t*-tests on individual variables are relatively unaffected by these factors.

Empirical error rates reported in the left side of table 3 show that given small sample sizes and normally distributed data, the empirical error rate for the chi-square test of the overall logit model is more sensitive to disparate response group sizes than the OLS F-test. Skewness increases the anticonservative bias of the test of the overall model and often results in nonconvergence when the sample is split into response groups of 10 and 90 percent. Disparate response group size also results in increased bias for the OLS F-test. The pattern of the bias for the OLS F-test, however, is erratic and would be difficult to adjust for heuristically.

The problem is model overfitting. With $n=50$ and 90 percent of the data in one group, there are only five data points in the other and four parameters to be estimated. Although this example is somewhat extreme, the last column of the appendix indicates that it is not unusual for accounting researchers to estimate parameters with ten or less observations in the smaller group.

Repeating the simulations with moderately correlated data ($\rho=0.50$), resulted in error rates that differed little from those on table 3. This indicates that the correlation of the predictor variables has little effect on empirical significance levels and hence that moderate multicollinearity does not seriously affect the calibration of test statistics for either the logit or the OLS model.

Implications of Results

The results reported in tables 1 to 3 show that logit test statistics are miscalibrated when sample sizes are small ($n=50$ or 100). *T*-tests of individual variables are conservatively biased, and the test of the overall logit model is anticonservatively biased. Skewness and disparate response group sizes exacerbate the problem. When data are skewed, as they often are in accounting research, OLS test statistics also are miscalibrated. Although statistically significant, the documented deviations for tests of the parameter estimates and the overall model are small and hence detectable only with a large number of replications.[12]

The question remains, then, as to whether the statistical differences noted are of practical significance. If practical significance is defined in terms of whether differences of the magnitude noted would have been likely to mislead researchers or readers of past studies to conclude relationships existed when they did not, the answer probably is no. This is because the anticonservatively biased test statistic, the chi-square test for the overall logit model, tends to be of only secondary interest in accounting studies. Examining the studies in the appendix, for example, shows that their focus was on hypotheses concerning individual parameter estimates and that the chi-square statistic was given with little comment. Further, in those studies with sample sizes of 50 or less, the reported chi-square statistics were so large relative to the critical value that cutting the statistic in half still would have resulted in a value that most researchers would conclude was significant. Because the t-tests of most of the studies would have been conservatively biased, readers can be even more confident that effects actually existed in studies in which null hypotheses concerning individual parameter estimates were rejected.

---

[12] This claim can be supported by an example from table 1, which reports an empirical error rate of 9.23 percent for a logit *t*-test with a nominal significance level of 10 percent and a sample size of 100. Converting to decimal notation, the difference between the nominal and empirical error rates is 0.0077. To have 95 percent confidence of detecting this deviation of $d=0.0077$ from the nominal significance level, it would be necessary to have 5,831 replications (determined using the standard statistical formula of $N=Z^2\alpha(1-\alpha)/d^2=(1.96)^2(0.1)(0.9)/(0.0077)^2=5{,}831$).

III. Other Issues

The empirical significance level of a test is not the only relevant issue in evaluating a statistical procedure. Such factors as interpretability of parameter estimates, predictive ability, and statistical power also are of interest. This section illustrates three problems—dependence of parameter estimates on the sampling range, impaired classification accuracy, and reduced power that arise when OLS is used to model dichotomous choices.

Effects of Sampling Range

Because the response variable is restricted to the interval between 0 and 1, a dichotomous choice model normally will be nonlinear. Whenever a linear function is used to model an inherently nonlinear relation, parameter estimates will be difficult (if not meaningless) to interpret. This is because there is no single slope for a nonlinear function. Rather, the rate of increase in the response variable varies, depending on the values of the predictor variables. Hence calculations using different ranges of the predictor variables will result in different "slopes" for the OLS linear model.

The effect of changes in the sampling range is illustrated by a small simulation of a hypothetical study attempting to predict bankruptcy on the basis of firm size. Only one predictor is used to simplify the presentation. Adding predictors would not affect the insights provided.

Ohlson's (1980) logit model for predicting bankruptcy within one year provides the starting point for the simulation. To represent the firm size variable, 50 random samples of $n=50$ and 100 observations[13] were generated from a uniform distribution[14] with a mean of 13.20 and a standard deviation of 1.58.[15] Because we wish to illustrate the problem of using OLS in a potentially nonlinear situation, binary data representing bankruptcy were generated, using a logit model with an intercept of 5.3724 and a slope of -0.407.[16] This procedure is then repeated (another 50 samples of $n=50$ and 100), with values of predictor variables now generated from a uniform distribution with the same

---

[13] In contrast to the previous simulation in which 10,000 replications were needed to isolate the factors causing miscalibration of logit and OLS test statistics, only 50 replications are necessary to demonstrate the sensitivity of OLS and the insensitivity of logit. Increasing the number of replications would increase the power of the test. However, this is unnecessary because of the extreme significance of the results reported in table 4.

[14] By standard statistical theory, data distributed uniformly on the interval from $\mu-\sqrt{3}\sigma$ to $\mu+\sqrt{3}\sigma$ will have mean μ and standard deviation σ. Use of the uniform distribution allows us to control directly the range of the predictor variable.

[15] This overall mean and standard deviation are obtained by combining Ohlson's reported values for bankrupt and nonbankrupt firms.

[16] The logit slope value of -0.407 is that reported by Ohlson. Ohlson's reported logit intercept reflects the inclusion of other variables in that model and hence is not relevant here. Instead the value of 5.3724 was selected to give firms of average size a bankruptcy probability of 50 percent plausible for a study with equal numbers of bankrupt and nonbankrupt firms.

Table 4
The Effect of Range of Observations on Logit and OLS Parameter Estimates[a] (True Logit $\beta = -.407$)

Mean $\hat{\beta}$ (standard deviation)

	Logit		OLS	
	N=50	N=100	N=50	N=100
Narrow Range[b]	−0.411 (0.204)	−0.427 (0.146)	−0.091 (0.040)	−0.097 (0.029)
Wide Range[c]	−0.400 (0.134)	−0.420 (0.092)	−0.078 (0.018)	−0.083 (0.012)
T-test (p-value)[d]	0.310 (0.760)	0.290 (0.792)	2.160 (0.033)	3.150 (<0.002)

[a] Based on 50 replications of each sample size.

[b] Mean and standard deviation for predictor variable = 13.20 and 1.58, respectively.

[c] Mean and standard deviation for predictor variable = 13.20 and 3.16, respectively.

[d] t-statistic and p-value for testing H_0: $\mu(\hat{\beta}$, WIDE$) = \mu(\hat{\beta}$, NARROW$)$, using the standard two-sample t-test.

mean (13.20) but with a standard deviation of 3.16, which is twice as large as before. Both logit and OLS models are fitted to each sample, and the resulting parameter estimates are compared.

Table 4 compares the mean parameter estimates for the logit and the OLS models. Panel A shows that the logit parameter estimates for the two sampling plans are the same to within sampling error. In contrast, OLS estimates in panel B are significantly different. Even though the true underlying relation between firm size and bankruptcy is the same for the two sampling plans, use of an OLS linear model produces significantly different estimates, which depend on the range of values from which the predictor variable is sampled. As the similarity of the OLS estimates for the two sample sizes indicates, increasing sample size does not eliminate the problem.

While this example has been simplified for ease of presentation, it does illustrate issues relevant to accounting research. It implies that two researchers could get two significantly different OLS estimates of the same true underlying relationship just because they used different data bases. In simulation, one sampling frame was arbitrarily established as twice as variable as the other. In practice, OLS parameter estimates would be affected by any change in the range of the predictors. The implications are that OLS estimates depend on the sampling frame and that care must be taken in interpreting them.

Predictive or Classificatory Accuracy: Auditors' Consistency Judgments

Classificatory ability is compared using the data base and model Stone and Ingram (1988) used to study auditors' consistency judgments. Real rather than simulated data are used to provide more concrete examples and to avoid the problem of specifying a "true" nonzero parameter vector that would be representative for accounting studies using logit.

The objective of the example is to compare the ability of logit and OLS to identify firms receiving consistency qualified opinions. To minimize the effect of sample selection and model overfitting on the results, the data base of 216 consistency judgments and related predictor variables was randomly split into an estimation sample of 108 and holdout sample of 108 observations. Parameter estimates and classification rates were aggregated over 250 repetitions of the splitting, estimation, and classification steps. This approach minimizes the effect of sample selection on test results. A similar approach was used by Elliott and Kennedy (1988) to compare multichotomous choice models.

Parameter estimates and descriptive statistics for the logit and OLS models used in the comparison are as follows:

Logit Model:

$$\ell n \frac{p}{1-p} = -1.80 + 1.33\,CURT + 599.06\,CINC/NS$$

(parameter estimates are medians for 250 replications of repeated splitting)

OLS Model: $\hat{y} = 0.27 + 0.20\,CURT + 59.29\,CINC/NS$

where:

$\hat{y}$ is 1 for changes resulting in consistency qualifications and 0 for changes not resulting in consistency qualifications;
$CURT$ is 1 for firms accounting for current period events under FASB Statement No. 88 and 0 otherwise; number coded 1=46, number coded 0=170;
$CINC$ is the after-tax income effect of the adoption of FASB Statement No. 87; and
NS is net sales.

$CINC/NS$: Mean=0.003; Std. dev.=0.004; Min.=0.00; Max.=0.03; and Skewness=2.73.

Table 5 compares the ability of logit and OLS to predict the correct classification for all firms and for firms receiving consistency qualified opinions. Panel A shows that the overall correct classification rate for logit (72.3 percent) and OLS (70.4 percent) is comparable. Panel B, however, shows that logit outperforms OLS in meeting the objective of the example—identifying firms receiving consistency qualified opinions. Logit correctly identifies 68 percent of the firms receiving consistency qualified opinions while OLS correctly identifies only 57.2 percent. The correct classification rate for unqualified opinions (not shown in the table) is 76.7 percent for logit and 83.7 percent for OLS. The null hypothesis that the classificatory accuracy of logit and OLS is equal can be rejected for each category of classification beyond the 0.001 level of significance using the McNemar test.[17] In 27,000 replications of this simulation, 4.2 percent of the predicted probabilities obtained by OLS were greater than one.[18]

---

[17] The McNemar test statistic is described in Conover (1980, 130).

[18] A practical solution to this problem is to set extreme predictions equal to 1 or 0. This, however, biases the probability estimates by predicting an occurrence with probability of 0 when it actually may occur or probability of 1 when it may not occur (Pindyck and Rubinfeld 1981). This is undesirable theoretically and results in measurement error when the probability estimate is used as a predictor variable in a subsequent model. The use of a logit probability estimate in a subsequent model is illustrated in Mittelstaedt (1989).

Table 5

Relative Predictive Ability of Logit and OLS (Estimation Sample = 108, Holdout Sample = 108; Replications with changing estimation and holdout sample = 250)

Panel A. Overall Prediction:

		Logit		
		Correct	*Incorrect*	*Total*
OLS	Correct	65.7%	4.7%	70.4%
	Incorrect	6.6	23.0	29.6
	Total	72.3%	27.7%	100.0%

Panel B. Prediction of Qualified Opinions:

		Logit		
		Correct	*Incorrect*	*Total*
OLS	Correct	56.4%	0.8%	57.2%
	Incorrect	11.6	31.2	42.8%
	Total	68.0%	32.0%	100.0%

Logit Model: $\ell n \frac{p}{1-p} = -1.80 + 1.33CURT + 599.06CINC/NS$

OLS Model: $\hat{y} = 0.27 + 0.20CURT + 59.29CINC/NS$

$CINC$ = after-tax income effect of the adoption of FASB Statement No. 87.

NS = net sales.

$CURT = 1$ for firms accounting for current period events under FASB Statement No. 88, 0 otherwise.

Power Comparison

The relative ability of OLS and logit to identify useful predictor variables also is of interest. Although detailed power comparisons would be complex and beyond the scope of this paper, our examination of relative classificatory ability provides some insight into the issue.[19] Both OLS and logit identified the *CINC/NS* variable as a meaningful predictor (significant at $\alpha = 0.05$) in all 250 replications. When a predictor variable is as significant as this one, almost any form of regression will be able to identify the relationship.

However, this will not be true when the relationship is weaker or harder to find, as often will be the case in accounting research. In the current study, logit and OLS differed in the frequency with which they identified the *CURT* predictor as significant. Logit identified it as significant in 61.2 percent of the replications, while OLS did so in only 47.2 percent of the replications. Using the McNemar test, the difference is statistically significant beyond the 0.001 level of significance, indicating that logit has greater power in this particular circumstance.[20] In general, logit can be expected to be more powerful whenever the relationship being modeled is nonlinear.

---

[19] The statistical power of logit procedures depends upon a variety of factors, including the true value of the logit parameters, the range and distribution of the predictors, and the sample size. Computations are complicated by the fact that there is no simple, closed-form formula for logit parameter estimates.

IV. Summary and Conclusions

Most researchers modeling dichotomous accounting choices assume their sample is "large enough" for the asymptotic properties of logit or probit to hold. However, a recent study (Noreen 1988) suggests that accounting sample sizes often are not "large enough" and that in these cases researchers may be better off using OLS. Our study provides benchmarks for assessing logit's sample size requirements when response group sizes are disparate and data are skewed and collinear, as they frequently are in accounting research. It also examines the sensitivity of logit and OLS parameter estimates to the range of observations sampled and compares the models' ability to predict auditors' consistency judgments.

Simulation results indicate that for scenarios comparable to those encountered in accounting research (four to six predictors and skewed data) sample sizes of 200 or more will be needed to ensure that logit test statistics will be properly calibrated (i.e., the nominal and empirical error rates will not be significantly different).

When sample sizes are smaller, logit test statistics will be modestly miscalibrated. However, much of the miscalibration actually will be attributable to skewness rather than to small sample size. *T*-tests for individual parameter estimates will reject the null hypothesis falsely less frequently than indicated by the nominal significance level. In contrast, the null hypothesis that the overall model has explanatory or predictive ability will be rejected falsely more frequently. Adherence to standard research procedures of considering the test of the overall model only jointly with tests of the individual coefficients will minimize the chance of being misled. As an additional safeguard, the test for the overall model could be performed at a somewhat reduced level of significance.

Using OLS rather than logit for sample sizes of 100 or less will result in test statistics that are somewhat better calibrated. However, the empirical error rates for these statistics also will differ from nominal rates when data are skewed and response group sizes disparate. The cost of using OLS will be parameter estimates that are sensitive to the range of observations sampled and reliance on a model with a functional form that may not reliably represent accounting choice. Whenever the functional form of the relationship is nonlinear, using OLS rather than logit can result in higher misclassification rates, a number of meaningless probability estimates, and less powerful tests of parameter estimates. Given these problems, we conclude that logit rather than OLS will continue to be the preferable method for modeling dichotomous accounting choices even when sample sizes technically are not "large enough."

---

[20] Simulation results reported earlier in the paper show that the probability of falsely rejecting the null hypothesis for the indicator variable is about the same for logit and OLS. Hence, we can be fairly confident that the difference noted is attributable to the power of the test.

Appendix

Accounting Studies Using Logit or Probit
Topic, Sample Size, Response Group Sizes, Number of Variables

Study	Model(s)	Total Sample Size	Response Group Sizes	No. of RHS Variables	Observations/ Variable in Smaller Group*
I. *Prediction*					
1. Dopuch et al. (1987)	P- WESML	564 470 440	218, 346 124, 346 94, 346	9	24 14 10
2. Palepu (1986)	L- WESML	419	163, 256	6–9	18
II. *Tests of Accounting Choice Explanations*					
3. Johnson and Ramanan (1988)	L	57	19, 38	6	3
4. Wong (1988a)	L	95	29, 66	3 or 5	6
5. Wong (1988b)	L	201	15, 186	5	3
6. Francis (1987)	L	800	213, 582 167, 582 162, 582 218, 582	2 or 3	71 56 54 73
7. El-Gazzar et al. (1986)	P	134 154	38, 96 41, 113	3 or 4	10 10
8. Ayres (1986)	L	232	103, 129	6	17
9. Zimmer (1986)	P	49	18, 31	3	6
10. Whittred (1987)	P	70	40, 30	4	8
III. *Tests of Explanations of Other Decisions*					
11. Francis and Wilson (1988)	P	78	21, 57	6–11	2
12. Shevlin (1987)	L	206	103, 103	8	13
13. Wilkerson (1987)	L	49	16, 33	3	5
14. Stone (1987)	P	216	58, 158	6	10
15. Casey et al. (1986)	P	113 72 41	56, 57 35, 37 21, 20	6	9 6 5
16. Pastena and Ruland (1986)	P	110	42, 68	4	11
17. Healy and Lys (1986)	L	110 105	94, 16 89, 16	4	24 22
18. Hamdallah and Ruland (1986)	P	80	40, 40	6	7

Appendix—*Continued*

IV. Evaluation of Judgment Consensus					
19. Morris and Nichols (1988) (individual parameter estimates were not reported)	L	42	12, 30	9	1
		31	15, 16		2
		29	17, 12		2
		49	17, 32		2
		33	10, 23		1
		41	20, 21		2
		69	21, 48		2
		40	15, 25		2

L = Logit regression.
P = Probit regression.
WESML = Weighted exogenous sample maximum likelihood.
[a] Rounded

References

Aldrich, J. H., and F. D. Nelson. 1984. *Linear Probability, Logit & Probit Models.* Quantitative Applications in the Social Sciences Series: No. 07-045. Beverly Hills: SAGE Publications.

Amemiya, T. 1981. Qualitative response models: A survey. *Journal of Economic Literature* 19 (December): 1483–536.

Ayres, F. L. 1986. Characteristics of firms electing early adoption of SFAS 52. *Journal of Accounting & Economics* 8 (June): 143–58.

Casey, C. J., V. E. McGee, and C. P. Stickney. 1986. Discriminating between reorganized and liquidated firms in bankruptcy. *The Accounting Review* 61 (April): 249–62.

Collins, R. A., and R. D. Green. 1982. Statistical methods for bankruptcy forecasting. *Journal of Economics and Business* 34 (October): 349–54.

Conover, W. J. 1980. *Practical Nonparametric Statistics.* 2nd ed. New York: John Wiley & Sons.

Dopuch, N., R. W. Holthausen, and R. W. Leftwich. 1987. Predicting audit qualifications with financial and market variables. *The Accounting Review* 62 (July): 431–54.

El-Gazzar, S., S. Lilien, and V. Pastena. 1986. Accounting for leases by lessees. *Journal of Accounting & Economics* 8 (October): 217–37.

Elliott, J. A., and D. B. Kennedy. 1988. Estimation and prediction of categorical models in accounting research. *Journal of Accounting Literature* 7: 202–42.

Francis, J. R. 1987. Lobbying against proposed accounting standards: The case of employers' pension accounting. *Journal of Accounting and Public Policy* 6 (Spring): 35–57.

———, and E. R. Wilson. 1988. Auditor changes: A joint test of theories relating to agency costs and auditor differentiation. *The Accounting Review* 63 (October): 663–82.

Frecka, T. J., and W. P. Hopwood. 1983. The effects of outliers on the cross-sectional distributional properties of financial ratios. *The Accounting Review* 58 (January): 115–28.

Hagerman, R. L., and M. E. Zmijewski. 1979. Some economic determinants of accounting policy choice. *Journal of Accounting & Economics* 1 (August): 141–61.

Hamdallah, A. E. S., and W. Ruland. 1986. The decision to terminate overfunded pension plans. *Journal of Accounting and Public Policy* 5 (Summer): 77–91.

Hanushek, E., and J. Jackson. 1977. *Statistical Methods for Social Scientists.* New York: Academic Press.

Healy, P., and T. Lys. 1986. Auditor changes following big eight mergers with nonbig eight audit firms. *Journal of Accounting and Public Policy* 5 (Winter): 251–65.

Hopwood, W., J. McKeown, and J. Mutchler. 1988. The sensitivity of financial distress prediction models to departures from normality. *Contemporary Accounting Research* 5 (Fall): 284–98.

Johnson, W. B., and R. Ramanan. 1988. Discretionary accounting changes from "successful efforts" to "full cost" methods. *The Accounting Review* 63 (January): 96–110.

McFadden, D. 1982. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics,* edited by P. Zarembka, 105–42. New York: Academic Press.

Maddala, G. S. 1983. *Limited-dependent and Qualitative Variables in Econometrics.* New York: Cambridge University Press.

Miller, R. G. 1986. *Beyond ANOVA: Basics of Applied Statistics.* New York: John Wiley & Sons.

Mittelstaedt, H. F. 1989. An empirical analysis of the factors underlying the decision to remove excess assets from overfunded pension plans. *Journal of Accounting & Economics* 11 (November): 399–418.

Morris, N. H., and W. D. Nichols. 1988. Consistency exceptions: Materiality judgments and audit firm structure. *The Accounting Review* 63 (April): 237–54.

Noreen, E. 1988. An empirical comparison of probit and OLS regression hypothesis tests. *Journal of Accounting Research* 26 (Spring): 119–33.

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18 (Spring): 109–31.

Palepu, K. G. 1986. Predicting takeover targets: A methodological and empirical analysis. *Journal of Accounting & Economics* 8 (March): 3–35.

Pastena, V., and W. Ruland. 1986. The merger/bankruptcy alternative. *The Accounting Review* 61 (April): 288–301.

Pindyck, R. S., and D. L. Rubinfeld. 1981. *Econometric Models & Economic Forecasts.* 2nd ed. New York: McGraw-Hill.

Shevlin, T. 1987. Taxes and off-balance-sheet financing: Research and development limited partnerships. *The Accounting Review* 62 (July): 480–509.

Stone, M. 1987. A financing explanation for overfunded pension plan terminations. *Journal of Accounting Research* 25 (Autumn): 317–26.

———, and R. Ingram. 1988. Auditors' use of quantitative decision rules in assessing the materiality of SFAS 87 adoption. Working paper, School of Accountancy, University of Alabama.

Whittred, G. 1987. The derived demand for consolidated financial reporting. *Journal of Accounting & Economics* 9 (December): 259–85.

Wilkerson, J. E., Jr. 1987. Selecting experimental and comparison samples for use in studies of auditor reporting decisions. *Journal of Accounting Research* 25 (Autumn): 161–67.

Wong, J. 1988a. Political costs and an intraperiod accounting choice for export tax credits. *Journal of Accounting & Economics* 10 (January): 37–51.

———. 1988b. Economic incentives for the voluntary disclosure of current cost financial statements. *Journal of Accounting & Economics* 10 (April): 151–67.

Zimmer, I. 1986. Accounting for interest by real estate developers. *Journal of Accounting & Economics* 8 (March): 37–51.

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 188–189

Activity Report for *The Accounting Review* for the Period June 1989–June 1990

THE manuscripts received and processed for *The Accounting Review* during my first year as Editor are classified by area, method, and action in exhibit 1. Panel A consists of the papers transferred to me from Bill Kinney, which included those with reviewers and those with authors for revision. Panel B includes the submissions made during my first year. In both panels, manuscripts are classified by subject matter (financial, auditing, cost-managerial, tax, governmental, not-for-profit, and others) and methodology (model building, quasi-experimental, experimental, and *a priori*). The actions taken are also cross-classified by function and method.

The turnaround time varied between three and six months; I have placed an arbitrary upper limit of six months even if I do not have all the reviews in. Six months is still quite long and I am concerned that a minority of reviewers are usually tardy. Innovative and well crafted manuscripts that have benefited by the critical comments of others have an easier time in the evaluation and review process.

The system of channeling many reviews through associate editors has worked reasonably well. I have also benefited in this respect by using several other faculty members as occasional (*ad hoc*) associate editors or reviewers as reported in the October 1990 issue. I am currently in the process of increasing the extent of delegated responsibilities to Associate Editors.

Finally, working with the publishing office at the AAA headquarters in Sarasota has gone quite well.

A. Rashad Abdel-khalik
Editor and Managing Editor

Exhibit 1
Types and Disposition of Manuscripts Through June 1990

Panel A. Manuscripts Inherited from Kinney's Period:

	M	Q	E	P	A	C	R
Financial	3	49	0	2	15	25	14
Auditing	7	9	21	1	5	21	12
Cost/Managerial	7	2	5	3	5	6	6
Tax	1	3	4	0	3	1	4
Governmental	1	1	0	0	0	2	0
Not-for-Profit	0	1	0	0	0	0	1
Other	1	5	8	7	4	4	13
A	6	17	8	1	32	—	—
C	11	30	12	6	—	59	—
R	3	23	18	6	—	—	50

Panel B. Manuscripts Submitted During June 1989 Through June 1990:

	M	Q	E	P	A	C	R
Financial	8	104	7	10	3	53	73
Auditing	12	18	44	2	3	32	41
Cost/Managerial	14	8	11	6	1	18	20
Tax	2	5	5	1	1	3	9
Governmental	0	3	0	0	0	1	2
Not-for-Profit	1	1	0	1	1	0	2
Other	8	15	21	23	2	13	52
A	0	2	4	5	11	—	—
C	17	68	32	3	—	120	—
R	28	84	52	35	—	—	199

Notes: M = Model building and analytical
Q = Quasi-experimental
E = Experimental
P = *A priori*
A = Accepted
C = Currently active
R = Rejected

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 190–201

American Accounting Association

Committees 1990–1991

EXECUTIVE COMMITTEE

President—Alvin A. Arens, Michigan State University
President-Elect—Arthur Wyatt, Arthur Andersen & Co.
Vice President—Wanda A. Wallace, Texas A&M University
Vice President—Mark A. Wolfson, Stanford University
Vice President—David A. Wilson, Ernst & Young
Secretary-Treasurer—Joseph J. Schultz, Jr., Arizona State University
Director of Education—Corine T. Norgaard, University of Connecticut
Director of Research—Nicholas Dopuch, Washington University
Director of Publications—Daniel W. Collins, University of Iowa
Past President—John K. Simmons, University of Florida

COUNCIL

Section and Group Representatives

Accounting, Behavior & Organizations Section—Kenneth Ferris, Southern Methodist University
Administrators of Accounting Programs Group—Jan R. Williams, University of Tennessee
Auditing Section—Andrew D. Bailey, Jr., University of Arizona and Carl S. Warren, University of Georgia
Government and Nonprofit Section—James M. Patton, University of Pittsburgh
International Accounting Section—Gary K. Meek, Oklahoma State University and Juan Rivera, University of Notre Dame
IS/MAS Section—Severin V. Grabski, Michigan State University
Management Accounting Section—Robert Capettini, San Diego State University and Donald K. Clancy, Texas Tech University
Public Interest Accounting Section—Tony Tinker, CUNY-Baruch College
Taxation (ATA) Section—Kenneth H. Heller, George Mason University and Edward J. Schnee, University of Alabama
Two-Year College Section—Helen Gerrard, Miami University

Regional Group Representatives

Mid-Atlantic Regional Group—John Cerepak, Fairleigh Dickinson University
Midwest Regional Group—Robert Holtfreter, Fort Hays State University and Paul Fischer, University of Wisconsin-Milwaukee
Northeast Regional Group—Samir Fahmy, St. John's University and Donald F. Arnold, Union College
Ohio Regional Group—Ray G. Stephens, Ohio State University
Southeast Regional Group—Robert West, University of South Florida and Robert W. Rouse, Auburn University
Southwest Regional Group—Clifford E. Hutton, University of Tulsa
Western Regional Group—J. Hal Reneau, Arizona State University and Jeffrey L. Harkins, University of Idaho

Members-at-Large

Matthew J. Anderson, Michigan State University
Victor Bernard, University of Michigan
Brenda Birkett, Southern University
Michael Gibbins, University of Alberta

REGIONAL VICE PRESIDENTS

Mid-Atlantic

John Cerepak, Fairleigh Dickinson University

Midwest

Paul Fischer, University of Wisconsin-Milwaukee

Northeast

Samir Fahmy, St. John's University

Ohio

Ray G. Stephens, Ohio State University

Southeast

Robert W. Rouse, Auburn University

Southwest

Clifford E. Hutton, University of Tulsa

Western

J. Hal Reneau, Arizona State University

SECTION OFFICERS

Accounting, Behavior & Organizations Section

Chairman: Howard Rockness, University of North Carolina
Vice Chairman (Academics): Mark Haskins, University of Virginia
Vice Chairman (Practice): Mike Gleason, Coopers & Lybrand
Secretary-Treasurer: Paul Munter, University of Miami

Administrators of Accounting Programs Group

President: Jan R. Williams, University of Tennessee
Vice President: James E. Smith, College of William & Mary
Secretary: Michael A. Diamond, UCLA
Treasurer: Penelope J. Yunker, Western Illinois University

Taxation (ATA) Section

President: Kenneth H. Heller, George Mason University
President-Elect: Edward J. Schnee, University of Alabama
Vice President: Caroline P. Strobel, University of South Carolina
Secretary-Treasurer: Debra W. Hopkins, University of Texas

Auditing Section

Chairperson: Carl S. Warren, University of Georgia
Vice Chairperson (Practice): Brent C. Inman, Coopers & Lybrand
Vice Chairperson (Academic): Theodore Mock, University of Southern California
Secretary-Treasurer: Zoe-Vonna Palmrose, University of Southern California

Gender Issues in Accounting

Chairperson: Paula B. Thomas, Middle Tennessee State University
Vice-Chairperson: Leslie Oakes, Rutgers University
Chairperson-Elect: Cheryl Lehman, Hofstra University
Secretary-Treasurer: Robert A. Nehmer, Ohio State University

Government and Nonprofit Section

Chairperson: James M. Patton, University of Pittsburgh
Chairperson-Elect: Jesse W. Hughes, Old Dominion University
Secretary: Earl Wilson, University of Missouri

Information Systems/Management Advisory Services Section

Chairperson: Severin Grabski, Michigan State University
Vice Chairperson (Practice): Thomas F. Farrell Coopers & Lybrand
Vice Chairperson (Academic), Chairperson-Elect: Casper E. Wiggins, Jr., Texas A&M University
Secretary-Treasurer: Anita S. Hollander, Florida State University

International Accounting Section

Chairman: Gary K. Meek, Oklahoma State University
Vice-Chairman (Academic): Juan Rivera, University of Notre Dame
Vice-Chairman (Practice): Alister Mason, Deloitte & Touche
Secretary: Kathleen Sinning, Western Michigan University
Treasurer: Maureen Berry, University of Illinois

Management Accounting Section

President: Robert Capettini, San Diego State University
Vice-President: Don Madden, University of Kentucky
Secretary-Treasurer: L. Gayle Rayburn, Memphis State University

Public Interest Accounting Section

Chair: Tony Tinker, CUNY-Baruch College
Vice Chair: Marilyn Neimark, CUNY-Baruch College
Secretary-Treasurer: Paula B. Thomas, Middle Tennessee State University
Chair-Elect: Jere R. Francis, University of Iowa

Two-Year College Section

Chairman: Helen Gerrard, Miami University
Vice Chairman: Billie Cunningham, Collin County Community College
Secretary: Patrick Reihing, Nassau Community College
Regional Program Chair: Beverly B. Terry, Central Piedmont Community College

ACCOUNTING ACCREDITATION COMMITTEE

Charges: 1. To monitor the AACSB review of accreditation and other developments in accreditation that would potentially impact accounting programs.

2. To inform the AAA Executive Committee of relevant accreditation issues and recommend policies or positions as appropriate.

Chairperson: Robert G. May, University of Texas–Austin
Kenneth Banet, Grant Thornton
David R. Campbell, Case Western Reserve University
Richard E. Flaherty, Arizona State University
John L. Kramer, University of Florida
Joseph A. McHugh, University of Massachusetts of Boston
Penelope J. Yunker, Western Illinois University

ACCOUNTING EDUCATION ADVISORY COMMITTEE

Charges: 1. To develop and issue position statements on matters relating to accounting education.

2. To develop and recommend to the Executive Committee educational policy for the Association.

3. To initiate, coordinate, and administer the activities of the Association in the field of accounting education, including accounting education research and continuing professional education. Effective liaison with regions and sections is an integral part of this responsibility.

4. To provide advice on accounting education matters to committees for the Doctoral Consortium, Doctoral Programs Conference, New Faculty Consortium, Senior Faculty Consortium, Trueblood Seminars, Professional Examinations, and other education-designated committees that may be established during the year.

Chairperson: Corine T. Norgaard, University of Connecticut
Elba F. Baskin, Arthur Andersen & Co.
Hugh A. Hoyt, Miami University
Paul A. Janell, Northeastern University
Daniel L. Jensen, Ohio State University
David N. Ricchiute, University of Notre Dame
Aurora Rubin, Deloitte & Touche
G. Fred Streuling, Brigham Young University
Milton F. Usry, University of West Florida
Thomas H. Williams, University of Wisconsin–Madison
Barbara Zarkowski, Bellevue Community College

ADMINISTRATIVE CHARGES EVALUATION COMMITTEE

Charges: 1. To study the need for and fairness of charging a fee for Association administrative services.

2. To recommend to the Executive Committee an equitable fee structure if an administrative fee is considered appropriate.

Chairperson: David B. Pearson, Ernst and Young
Robert E. Holtfreter, Fort Hayes State University
Penny R. Marquette, University of Akron
Shane R. Moriarity, University of Oklahoma
Hadley P. Schaefer, University of Florida
Michael Sutton, Deloite & Touche

ASSOCIATION STAFFING EVALUATION COMMITTEE

Charges: 1. To identify Association activities currently performed by volunteers that could be more effectively performed by paid permanent staff.

2. To identify Association activities currently performed by paid permanent staff that could be more effectively contracted or performed by volunteers.

3. To recommend to the Executive Committee changes, if any, in the responsibilities of paid permanent staff.

4. To plan a strategy for the eventual succession of the Association Executive Director.

5. To recommend to the Executive Committee a financing plan for all recommendations.

Chairperson: Roger H. Hermanson, Georgia State University
Jack C. Gray, University of Minnesota
Thomas E. McKee, East Tennessee State University
Deborah H. Turner, Georgia Institute of Technology
Norman R. Walker, Price Waterhouse

AWARDS EVALUATION COMMITTEE

Charges: 1. To recommend criteria the Executive Committee should use in deciding the nature of awards the Association will offer.

2. To review and evaluate existing Association awards to determine whether policies and procedures for all awards are consistent with the criteria.

3. To recommend additional awards, the elimination of existing ones, and changes in the policies and procedures for selecting award recipients.

Chairperson: Yuji Ijiri, Carnegie Mellon University
Thomas R. Dyckman, Cornell University
Thomas J. Frecka, University of Illinois
Rita E. Hull, Virginia Commonwealth University
Philip H. Siegel, University of Houston-Downtown

BYLAWS COMMITTEE

Charge: To review the Association's Bylaws and recommend appropriate revisions to the Executive Committee.

Chairperson: Charles G. Carpenter, Miami University
Harold Q. Langenderfer, University of North Carolina
Shyam Sunder, Carnegie Mellon University
Nancy O. Tang, Portland State University

CARTER SCHOLARSHIPS COMMITTEE

Charges: 1. To collect and evaluate applications for the Carter Foundation Student Scholarships.

2. To recommend amounts of stipends.

3. To rank applications and to transmit the ranking to the trustees of the Foundation.

Chairperson: Eldon C. Lewis, California State University–San Bernadino
Geraldine F. Dominiak, Texas Christian University
N. Allen Ford, University of Kansas

COMPETITIVE MANUSCRIPT AWARD COMMITTEE

Charge: To select one, two, or, at most, three recipient(s) of the Competitive Manuscript Award using criteria approved by the Executive Committee.

Chairperson: William S. Waller, University of Arizona
Timothy B. Bell, KPMG Peat Marwick
Lawrence D. Brown, State University of New York at Buffalo
David C. Burgstahler, University of Washington
Judy D. Rayburn, University of Minnesota
Charles W. Swenson, University of Southern California
H. Thomas Johnson, Portland State University

CONFERENCE, CONSORTIA, AND SEMINAR EVALUATION COMMITTEE

Charges: 1. To recommend criteria the Executive Committee should use in deciding the nature of conferences, consortia, and seminars the Association will offer.

2. To evaluate existing Association conferences, consortia, and seminars to determine whether their purposes and content are consistent with the criteria.

3. To evaluate existing Association conferences, consortia, and seminars to determine whether policies and procedures for invitations, fees, funding, responsibilities of committees, and other matters are equitable and consistently applied.

4. To recommend changes in policies and procedures for their administration.

Chairperson: Frank R. Rayburn, University of Alabama at Birmingham
Ronald C. Mannino, University of Massachusetts
Barbara D. Merino, University of North Texas
Loren A. Nikolai, University of Missouri at Columbia
Larry E. Rittenberg, University of Wisconsin–Madison
Larry Scott, Price Waterhouse
Ronald V. Hartley, Bowling Green State University

CONTINUING PROFESSIONAL EDUCATION EVALUATION COMMITTEE

Charges: 1. To develop criteria for evaluating successful continuing professional education programs sponsored by the Association.

2. To evaluate existing continuing education programs using the developed criteria.

3. To evaluate whether it is practical and desirable to finance other Association activities by increasing fees or expanding continuing professional education programs.

4. To evaluate the need for paid professional assistance in marketing and administering the Association's continuing professional education effort.

5. To coordinate all activities and recommendations with the Accounting Education Advisory Committee.

Chairperson: Donald L. Madden, University of Kentucky
Brent Inman, Coopers & Lybrand
Malcolm H. Lathan, University of Virginia
Richard J. Murdock, Ohio State University
Thomas Nessinger, Arthur Andersen & Co.–Chicago
Mary Ellen Oliverio, Pace University
Robert W. Rouse, Auburn University
William T. Wrege, Ball State University

CORPORATE ACCOUNTING POLICY SEMINAR COMMITTEE

Charges: 1. To plan the 1991 Corporate Accounting Policy Seminar in accordance with established AAA policies.

2. To arrange to have the seminar conducted and administered.

3. To report on the results and effectiveness of the seminar and to recommend changes by November 1, 1991.

Chairperson: James Don Edwards, University of Georgia
Carolyn M. Callahan, University of Massachusetts
Roland E. Dukes, University of Washington
Eugene Flegm, General Motors Corporation
William Ihlanfeldt, Shell Oil Company
Edward J. Joyce, University of Minnesota
Paul B. W. Miller, University of Colorado–Colorado Springs
Phyllis Minnott, Eli Lilly & Company

DELOITTE & TOUCHE WILDMAN MEDAL AWARD COMMITTEE

Charge: To administer the John R. Wildman Medal award program in accordance with the terms of the grant from the Deloitte & Touche Foundation.

Chairperson: Andrew D. Bailey, Jr., University of Arizona
Eugene E. Comiskey, Georgia Institute of Technology
Jack E. Kiger, University of Tennessee
Wai P. Lam, University of Windsor
Cheryl R. Lehman, Hofstra University
Joseph G. San Miguel, Naval Postgraduate School
Lester M. Sussman, Deloitte & Touche

DOCTORAL CONSORTIUM COMMITTEE

Charges: 1. To plan the 1991 Doctoral Consortium in accordance with established AAA policies.

2. To arrange to have the consortium conducted and administered.

3. To report on the results and effectiveness of the consortium and to recommend changes by August 1, 1991.

Chairperson: Gerald L. Salamon, Indiana University
Dan S. Dhaliwal, University of Arizona
Susan F. Haka, Michigan State University
James J. Jiambalvo, University of Washington
Robert P. Magee, Northwestern University

DOCTORAL FELLOWSHIP COMMITTEE

Charge: To select the 1990–1991 recipients of American Accounting Association fellowship grants using the criteria approved by the Executive Committee and to determine the amounts of these grants.

Chairperson: A. Faye Borthick, University of Tennessee
Stanley F. Biggs, University of Connecticut
Constantine Konstans, Memphis State University

DOCTORAL PROGRAMS CONFERENCE COMMITTEE

Charges: 1. To finalize the plans for the 1991 Doctoral Programs Conference in accordance with established AAA policies.

2. To arrange to have the conference conducted and administered.

3. To report on the results and effectiveness of the conference and to recommend changes by June 1, 1991.

Chairperson: Charles H. Smith, Penn State University
Michael J. Gift, University of South Florida
W. Robert Knechel, University of Florida
J. Hal Reneau, Arizona State University

FINANCE COMMITTEE

Charges: 1. To prepare financial goals and strategies for the Association.

2. To evaluate additional or alternative ways to finance existing Association activities.

3. To work with the 1990–1991 Present-Elect in preparing the 1991–1992 budget.

4. To recommend how AAA funds should be invested in accordance with established Executive Committee policies and procedures.

Chairperson: Joseph J. Schultz, Arizona State University
William F. Bentz, Ohio State University
Robert E. Jensen, Trinity University
Konrad W. Kubin, Virginia Polytechnic Institute and State University
Gary E. White, Texas Tech University
David A. Wilson, Ernst & Young

FINANCIAL ACCOUNTING STANDARDS COMMITTEE

Charges: 1. To evaluate selected discussion memoranda and exposure drafts related to financial accounting and reporting in the private sector as they are released by the FASB and to respond to the FASB in writing and by appearing at selected public hearings.

2. To be cognizant of emerging issues related to financial accounting and reporting in the private sector through materials issued by the Emerging Issues Task Force and other relevant sources with evaluation, where appropriate, of the significance and potential implications of such issues.

3. To meet with the FASB, normally on an annual basis.

4. To explore the publication of selected responses by the committee in *Horizons* or other appropriate outlets.

5. To consider and promote ways to increase the academic input to the standard-setting process.

6. To provide input to the Financial Reporting Issues Conference Committee on the appropriate content and structure of the 1991 conference.

Chairperson: Jerry J. Weygandt, University of Wisconsin–Madison
William A. Collins, North Carolina at Greensboro
Michael Crooch, Arthur Andersen & Co.
Loyd C. Heath, University of Washington
William W. Holder, University of Southern California
Charles L. McDonald, University of Florida
Lawrence Revsine, Northwestern University
D. Gerald Searfoss, Deloitte & Touche
Lawrence A. Tomassini, University of Illinois

FINANCIAL REPORTING ISSUES CONFERENCE COMMITTEE

Charges: 1. To plan the 1991 Financial Reporting Issues Conference in accordance with established AAA policies.

2. To seek the advice of the Financial Accounting Standards Committee on the content and structure of the 1991 conference.

3. To arrange to have the conference conducted and administered.

4. To prepare proceedings of the conference for distribution by the Association.

5. To report on the results and effectiveness of the conference and to recommend changes by November 1, 1991.

Chairperson: William W. Holder, University of Southern California
Robert K. Eskew, Purdue University
Charles L. McDonald, University of Florida
Maureen F. McNichols, Stanford University
D. Gerald Searfoss, Deloitte & Touche

HISTORY PROJECT ADVISORY COMMITTEE

Charge: To supervise the execution of the history project with the objective of completing the monograph within the project budget in time for the August 1991 AAA Annual Meeting.

Chairperson: Thomas J. Burns, Ohio State University
Thomas J. Nessinger, Arthur Andersen & Co.
Stephen A. Zeff, Rice University

INTERNATIONAL FACULTY EXCHANGE COMMITTEE

Charges: 1. To select the professors to be involved in the program for 1991–1992.

2. To assist the Executive Director with announcements, arrangements, and follow-up evaluations for the respective visits.

Chairperson: Lauren Kelly, University of Washington
Kathleen R. Bindon, University of Alabama
Nabil M. A. H. Hassan, Wright State University
Marcos F. Massoud, Claremont McKenna College
Gary K. Meek, Oklahoma State University
Andre Zund, University of St. Gallen

MEMBERSHIP AND SUBSCRIPTIONS COMMITTEE

Charges: 1. To increase the quantity and quality of the membership of the Association.

2. To increase the quantity of the subscriptions to all Association journals, but especially *Accounting Horizons* and *Issues in Accounting Education.*

Chairperson: Richard E. Baker, Northern Illinois University
Joseph H. Anthony, Michigan State University
Mary S. Doucet, University of Georgia
James B. Edwards, University of South Carolina
Michael Egan, Ernst & Young
Jeffrey L. Harkins, University of Idaho
James R. Hasselback, Florida State University
Bernie Milano, KPMG
Linda M. Nichols, Texas Tech University
Blaine A. Ritts, Bowling Green State University
Bill N. Schwartz, Virginia Commonwealth University
Beverly B. Terry, Central Piedmont Community College
Jimmy Yang-Tzong Tsay, National Taiwan University

MINORITY FACULTY DEVELOPMENT COMMITTEE

Charges: 1. To summarize the recommendations made by the Minority Faculty Development Committees for the past five years and evaluate the impact these recommendations are having on the Association's activities.

2. To explore additional ways in which potential minority candidates can be encouraged to apply to accounting doctoral programs and to pursue professional careers in accounting education.

3. To explore ways in which to encourage the development of minority faculty members.

4. To examine and monitor minority recruitment programs of other professional organizations, such as the AICPA.

Chairperson: Margaret Hicks, Howard University
Joyce H. Holley, Texas A&M University
Johnnie A. Mapp, Norfolk State University
Benjamin Newhouse, Tuskegee University
Wilbur I. Smith, Florida A&M University
Jenice P. Stewart, University of Missouri at Columbia

MISSION STATEMENT COMMITTEE

Charges: 1. To evaluate the stated purposes and objectives of the Association as they are currently included in the Bylaws.

2. To develop and recommend to the Executive Committee a mission statement that includes but is not limited to the following areas:

a. Response to the changing nature of accounting education.
b. Role of teaching, research, and publications in the Association's activities.
c. Association involvement in international activities.
d. Role of practitioners and their organizations in relation to the Association.
e. Role of the Association in responding to activities of other organizations that influence accounting education, research, and standard-setting.

Chairperson: Gerhard G. Mueller, University of Washington
Peter E. Battelle, University of Vermont
William H. Beaver, Stanford University
Robert K. Elliott, KPMG Peat Marwick Main & Co.
Donald J. Kirk, Columbia University
Jack L. Krogstad, Creighton University
Zoe-Vonna Palmrose, University of Southern California

NEW FACULTY CONSORTIUM COMMITTEE

Charges: 1. To plan the 1991 New Faculty Consortium in accordance with established AAA policies.

2. To arrange to have the consortium conducted and administered.

3. To report on the results and effectiveness of the consortium and to recommend changes by March 1, 1991.

Chairperson: Helen Gernon, University of Oregon
Michael A. Diamond, University of Southern California
Silvia A. Madeo, University of Missouri–St. Louis
Dennis Reigel, Arthur Andersen & Co.
Mikel G. Tiller, Indiana University

NOMINATIONS COMMITTEE

Charge: To select a list of nominees to AAA offices for election by the membership in August 1991.

Chairperson: William H. Beaver, Stanford University
Robert J. Capettini, San Diego State University
Donald K. Clancy, Texas Tech University
Jeffrey L. Harkins, University of Idaho
Gerhard G. Mueller, University of Washington
Joanne Rockness, North Carolina State University
John K. Simmons, University of Florida

OUTSTANDING ACCOUNTING EDUCATOR AWARD COMMITTEE

Charge: To select one or, at most, two recipients of the Outstanding Educator Award using the criteria approved by the Executive Committee.

Chairperson: James R. Boatsman, Arizona State University
James L. Chan, University of Illinois at Chicago
Michael H. Granof, University of Texas at Austin
Rosanne M. Mohr, Portland State University
James M. Patell, Stanford University
G. Fred Streuling, Brigham Young University

PROFESSIONAL EXAMINATIONS COMMITTEE

Charges: 1. To monitor activities of the AICPA and NASBA and evaluate the implications for the CPA examination and accounting education.

2. To identify activities the Association should consider undertaking that deal with the effect of professional examinations on accounting education.

3. To coordinate all activities and recommendations with the Accounting Education Advisory Committee.

Chairperson: Alan Reinstein, Wayne State University
Joyce S. Allen, Xavier University
James D. Blum, American Institute of Certified Public Accountants
William R. Cron, Central Michigan University
David R. L. Gabhart, Bentley College
John M. Hassell, University of Texas at Arlington
C. Jevons Lee, Tulane University
Kenneth A. Smith, Idaho State University

PROFESSIONALISM AND ETHICS SEMINAR COMMITTEE

Charges: 1. To plan the 1991 Professionalism and Ethics Seminar in accordance with established AAA policies.

2. To arrange to have the seminar conducted and administered.

3. To report on the results and effectiveness of the seminar and to recommend changes by July 1, 1991.

Chairperson: Joanne Rockness, North Carolina State University
Mary Beth Armstrong, California Polytechnic State University
James C. Gaa, McMaster University
Stephen E. Loeb, University of Maryland
William May, University of Southern California
Robert J. Stack, University of Virginia
Richard P. Weber, Michigan State University

PROGRAM ADVISORY COMMITTEE

Charge: To assist the President and the Executive Director in developing the technical program for the 1991 Annual Meeting.

Sections Liaison Representatives

Accounting, Behavior, and Organizations Section—S. Mark Young, University of Colorado at Boulder
Administrators of Accounting Programs Group—James E. Smith, College of William and Mary
American Taxation Association—Michael L. Roberts, University of Alabama
Auditing—Jane F. Mutchler, University of Arizona
Gender Issues in Accounting (Provisional) Section—Karen L. Hooks, University of South Florida
Government and Nonprofit—Robert W. Parry, Jr., Indiana University
Information Systems/Management Advisory Services: Carol E. Brown, Oregon State University
International Accounting Section—James A. Schweikart, University of Richmond
Management Accounting—Fred H. Jacobs, Michigan State University
Public Interest Section—Jesse Dillard, Ohio State University
Two-Year College—Billie Cunningham, Collin County Community College

Education Subset

Chairperson: Frederick L. Neumann, University of Illinois
Vice-Chairperson: James C. McKeown, Penn State University
Richard P. Brief, New York University
C. Dwayne Dowell, Abilene Christian University
Terry J. Engle, University of South Florida
Hartwell C. Herring III, University of Tennessee
Kenneth A. Merchant, Harvard University
Jean Wyer, Coopers & Lybrand

Academic Subset

Stanley Baiman, Carnegie Mellon University
William S. Hopwood, University of Houston
Marian Powers, University of Illinois at Chicago
Thomas F. Schaefer, Florida State University
Michael D. Shields, San Diego State University

Applied Subset

Richard Boland, Case Western Reserve University
Douglas R. Carmichael, CUNY-Baruch College
Henry R. Jaenicke, Drexel University
Wayne J. Morse, Clarkson University
Felix Pomeranz, Florida International University
Karin M. Skadden, George Mason University

PUBLICATIONS COMMITTEE

Charges: 1. To monitor all aspects of the Association's publications program.

2. To study and evaluate additional or alternative ways to finance Association journals.

3. To report periodically to the President and Executive Committee on all matters of concern in the field of publications.

4. To recommend candidates for editor or editor-elect positions as appropriate.

5. To recommend AAA publications policies to the Executive Committee as deemed desirable.

Chairperson: Daniel W. Collins, University of Iowa
Robert H. Ashton, Duke University
Jane O. Burns, Texas Technology University
Chee W. Chow, San Diego State University
William L. Felix, Jr., University of Arizona
William R. Kinney, Jr., University of Texas at Austin
John Willingham, KPMG Peat Marwick Main & Co.

RESEARCH ADVISORY COMMITTEE

Charges: 1. To monitor the accounting research activities of other accounting associations, e.g., AICPA, FEI, IIA, and NAA.

2. To advise the Director of Research in the development and administration of the Association's program in accounting research.

3. To suggest ways to encourage research in accounting, especially applied research dealing with all accounting functions in enterprise settings.

Chairperson: Nicholas Dopuch, Washington University
Rick Antle, Yale University
Ronald A. Dye, Northwestern University
Dan A. Simunic, University of British Columbia
Abbie Smith, University of Chicago
Ira Solomon, University of Illinois

SCREENING COMMITTEE: NOTABLE CONTRIBUTION TO ACCOUNTING LITERATURE AWARD

Charge: To identify works of exceptional merit from published accounting books and articles that meet the guidelines for the "Notable Contributions" awards established by the Executive Committee.

Chairperson: Barry L. Lewis, University of Colorado at Boulder
John C. Anderson, Syracuse University
Andrew H. Barnett, San Diego State University
Robert M. Bushman, University of Chicago
John Cumming, Miami University
John H. Evans III, University of Pittsburgh
G. William Glezen, University of Arkansas
Lester E. Heitger, Indiana University
Harvey S. Hendrickson, Florida International University
Rhoda C. Icerman, Florida State University
Larry N. Killough, Virginia Polytechnic Institute and State University
Suzanne M. Luttman, University of Colorado at Boulder
Janet A. Meade, University of Houston
Kathryn M. Means, Florida Atlantic University
Hugo Nurnberg, CUNY–Baruch College
Patricia C. O'Brien, University of Michigan
Daniel Pearl, University of Lowell
Graeme W. Rankine, University of Arizona
Sue Ravenscroft, Eastern Michigan University
Jane L. Reimers, Florida State University
Byung-Tak Ro, Purdue University
Avi Rushinik, University of Miami
Jeffrey W. Schatzberg, University of Arizona
Michael T. Stein, Rutgers University–New Brunswick
Clyde P. Stickney, Dartmouth College
Jacob K. Thomas, Columbia University
Mary C. Tichich, The American University
Richard M. Tubbs, University of Iowa
Kiran Verma, Harvard University
Paul F. Williams, North Carolina State University
Alan J. Winters, University of South Carolina

SECURITIES AND EXCHANGE COMMISSION LIAISON COMMITTEE

Charges: 1. To prepare and transmit a response to the staff of the Securities and Exchange Commission on proposals soliciting views from the public concerning auditing and financial accounting and reporting.

2. To make recommendations and to develop program(s) that will assist teaching of and conducting research in topics concerning the SEC and its activities.

3. To conduct activities as appropriate to assist communications and interactions between the SEC and the membership of the Association.

Chairperson: Mary S. Stone, University of Alabama
Charles W. Mulford, Georgia Institute of Technology
David B. Smith, Claremont McKenna College
David E. Stout, Villanova University
Thomas R. Weirich, Central Michigan University

SELECTION COMMITTEE: NOTABLE CONTRIBUTION TO ACCOUNTING LITERATURE AWARD (JOINT COMMITTEE WITH THE AICPA)

Charge: To select the recipient(s) of the award from those books and articles identified by the Screening Committee as potentially notable accounting literature contributions.

Chairperson: Wilfred C. Uecker, Rice University
Julie H. Collins, University of North Carolina
Michael Gibbins, University of Alberta
Eugene A. Imhoff, Jr., University of Michigan
Joshua Ronen, New York University
Carl S. Warren, University of Georgia

SENIOR FACULTY CONSORTIUM COMMITTEE

Charges: 1. To plan the 1991 Senior Faculty Consortium in accordance with established AAA policies.
2. To arrange to have the consortium conducted and administered.
3. To report on the results and effectiveness of the consortium and to recommend changes by August 1, 1991.

Chairperson: Jay M. Smith, Jr., Brigham Young University
Thomas P. Howard, Arizona State University—West Campus
Kurt J. Pany, Arizona State University
Ann B. Pushkin, West Virginia University
Gerald Smith, University of Northern Iowa

SEVENTH INTERNATIONAL CONFERENCE ON ACCOUNTING EDUCATION (1992) PLANNING COMMITTEE

Charge: To continue planning for the conference.

Chairperson: Belverd E. Needles, Jr., DePaul University
Thomas A. Gavin, University of Tennessee at Chattanooga
Sidney J. Gray, Glasgow University
W. Morley Lemon, University of Waterloo
William Markell, University of Delaware
Seigo Nakajima, Ferris Jogakuin Women's College
Walter F. O'Connor, Fordham University
Norlin G. Rueschhoff, University of Notre Dame
Murray C. Wells, University of Sydney

TEACHING/CURRICULUM DEVELOPMENT AWARD COMMITTEE

Charge: To select recipient(s) of the Outstanding Teaching/Curriculum Development Award using the criteria approved by the Executive Committee.

Chairperson: Harold M. Sollenberger, Michigan State University
W. Steve Albrecht, Brigham Young University
Richard J. Asebrook, University of Massachusetts
W. Thomas Harrison, Baylor University
George W. Krull, Jr., Grant Thornton
Karen V. Pincus, University of Southern California

TRUEBLOOD SEMINARS COMMITTEE

Charges: 1. To assist Deloitte & Touche in the conduct of the Trueblood Seminars.
2. To report on the results and effectiveness of the seminars and to recommend changes by August 1, 1991.

Chairperson: Karen L. Hooks, University of South Florida
Mary E. Barth, Harvard University
R. D. Nair, University of Wisconsin–Madison
Larry Ponemon, SUNY at Albany

THE ACCOUNTING REVIEW
Vol. 66, No. 1
January 1991
pp. 202–210

BOOK REVIEWS

Shane R. Moriarity, Editor

Editor's Note: Two copies of books for review should be sent to Professor Shane Moriarity, School of Accounting, University of Oklahoma, Norman, OK 73019. The policy of the *Review* is to publish only those reviews solicited by the Book Review Editor. Unsolicited reviews will not be accepted.

JODY BLAZEK, *Tax and Financial Planning for Tax-Exempt Organizations: Forms, Checklists, Procedures* (New York: John Wiley & Sons, Inc., 1990, pp. xiv, 446, $85.00).

This book is a practical guide to establishing and maintaining tax-exempt status and fiscally managing an exempt organization. The objective of the book is to provide guidelines for compliance and effective management of nonprofit organizations. Types of organizations that *can* qualify and those that *cannot* are compared and contrasted. A compendium of checklists, client memoranda, and compliance tools; this book was specially developed as a companion to *The Law of Tax-Exempt Organizations*, by Bruce Hopkins, Esq., and offers practitioners and managers valuable step-by-step guidance in obtaining and safeguarding the tax-exempt status of an organization.

One major strength of the work lies in the well-organized and useful checklists designed for use by practitioners, organization managers, and board members or trustees involved in nonprofit organizations. While being accurate from a technical standpoint, the materials are understandable by individuals not knowledgeable about the subject. The inclusion of actual tax forms and instructions for maintaining, on a day-to-day basis, the necessary information for filing such returns adds to the usefulness of this book. A glossary of financial terms and abbreviations is especially valuable to the nonfinancial manager of such an organization.

My experience with nonprofit organizations supports the need for this type of book. Usually, professionals and/or volunteers in the nonprofit endeavor prefer to focus on the goals and services of the organization with minimum time devoted to financial and tax matters. This book will enable these individuals to do just that. I would strongly recommend this book for use by managers in these organizations since the material is presented in a well-organized, concise, understandable format. This book can serve as a ready reference for basic questions regarding financial aspects of the management of the organization, thus saving time and the expense of outside consultants.

Chapter 1 begins with ten questions to ask before starting an organization. It provides a clear and concise understanding of the distinctive nature of nonprofit organizations. The last half of Chapter 1 details the requirements for exemption of the various Section 501(c) categories and ends with a chart comparing them.

The next chapter provides a blueprint for seeking approval for exempt status from the Internal Revenue Service, including specific instructions for answering each question on Forms 1023 and 1024. Examples of these forms are included which adds to the discussion in the chapter.

Annual tax compliance checklists are the major subject of Chapter 3. These checklists will help verify quickly and accurately whether an exempt organization still qualifies. Suggestions for completion of Forms 990, with filled in examples, are included. The section on "weathering an IRS examination" could prove invaluable to managers in the throes of dealing with the IRS.

Chapter 4 describes which organizations are classified as private foundations and the special rules applicable to them. The dizzying labyrinth of excise taxes, their rates, and to whom they apply can be made simpler using the material in this chapter.

The book ends with a discussion of several aspects of financial management and planning for nonprofits. Charts and checklists of duties of the board members, organization treasurer, and inside versus outside accountants, provide operational guidelines. Sample financial statements, budgeting and internal control guides, and a financial planning checklist are presented.

A valuable index and bibliography is included. As stated earlier, this book is well organized and presents a considerable amount of useful data. Blazek does an excellent job of presenting checklists and information understandable by individuals not knowledgeable about nonprofit organizations. This book should be added to the reference materials in every nonprofit organization.

DARLENE K. AHNBERG
Associate Professor of Accounting

The University of Akron

M. BROMWICH and A. BHIMANI, *Management Accounting: Evolution not Revolution* (London: The Chartered Institute of Management Accountants, pp. vii, 112, £9.96, paper).

This book is based on a research study commissioned by the Chartered Institute of Management Accountants to investigate the relationship between the changes in the manufacturing environment and their impact on management accounting. Specifically, the purpose of the report was to identify "the criticisms being made, the way in which management accounting techniques already exist to meet them, and the findings of research in this field and any gaps that might initiate further research" (p. 1). The book identifies some challenges and opportunities related to management accounting in the new manufacturing environment and suggests some possible actions for the benefit of practicing management accountants.

The authors initially identify investment appraisal, performance evaluation, and cost management as areas that need to be reassessed in light of the changes being made in the manufacturing environment. This is followed by a concise description of Materials Requirements Planning, Just-in-Time Systems, Numerical Control Machines, Computer-Aided Design and Computer-Aided Manufacturing, and Flexible Manufacturing Systems. The capabilities of these production systems and advanced manufacturing technologies (AMT) are discussed and the major accounting problems related to them are introduced. Schematic drawings and/or photos of the AMTs might have helped the uninformed reader to visualize these technologies.

In the chapter, "Japanese Management Accounting and AMT," attention is focused on the management accounting practices of leading and innovative Japanese firms. However, the different accounting practices discussed are derived from empirical studies that do not relate the accounting practices to only companies with different forms of AMTs. Therefore, the linkage between Japanese management accounting and AMT is weak. The authors suggest correctly that the close link between corporate strategy and accounting policies and systems found in Japanese firms may be useful in Western firms. In addition, the need for Western firms to expand their emphasis on long-term results and nonfinancial quantitative and qualitative measures is recognized.

In discussing the "Crisis" in management accounting from the U.S. perspective, the authors focus on areas such as the primacy of financial reporting and the components of product cost and cost management, as well as the need for better investment appraisal, product costing, and performance measures for AMT. They propose to evaluate AMT investments within a strategic framework in which both quantifiable and qualitative factors are considered. The idea is appealing but an example would have helped to assess the merits of their proposal. An example was given showing how conventional product costing distorts product costs and how activity-based costing (ABC) assigns costs more equitably in AMT environments. The example inappropriately included direct material cost and direct labor cost (which was quite high) in overhead cost and two errors in the tables and discussion which followed made the example somewhat hard to follow. The discussion of the merits and limitations of ABC was good. One of the authors' conclusions in this section is that many problems of U.S. companies can be solved by changing ingrained business management practices rather than by changing accounting practices.

The last two chapters deal with AMT and the current and future status of management accounting in the U.K. Numerous empirical studies in the U.K. are summarized and the implications for management accounting practitioners emphasized. Some of the implications for management accountants include the need to focus on longer-term results, use more nonfinancial performance measures, adopt a strategic perspective, concentrate on supplying relevant information, and communicate better with information users. The authors stress that management accountants need to consider fully the operational activities that are organization specific. Therefore, implementing new accounting techniques devised in the U.S. or Japan for the changing manufacturing environment may not be suitable for British firms because of their unique characteristics and the specific economic, political, and social conditions in their environment.

The authors conclude that "no general crisis has been identified within the management accounting profession in the U.K. vis-à-vis the changing manufacturing environment and therefore no radical reforms are recommended at this stage" (p. 6–7). They advocate a greater understanding of the important factors related to the new manufacturing environment and the information needs of managers before adopting precipitous changes in management accounting systems. The authors do recommend that the use of certain techniques such as ABC and target pricing be implemented and further developed. Experimentation and innovation in management accounting is encouraged in individual firms as a means of determining how management accounting systems must change to provide relevant information in the new manufacturing environment.

This book provides an excellent summary of the changes in the new manufacturing environment and the implications of these changes for management accountants. A comprehensive bibliography provides a wealth of sources for anyone who wants to research any area further. While the primary emphasis is on the U.K., the book is broadly focused so as to be of interest to readers in the U.S., Japan, and other countries. I recommend it to any reader who wants a concise, comprehensive treatment of the new issues in management accounting related to AMT.

JAMES A. HENDRICKS
Square D Professor of Accountancy
Northern Illinois University

YASUHIRO MONDEN and MICHIHARU SAKURAI, Editors, *Japanese Management Accounting: A World Class Approach to Profit Management* (Cambridge, MA: Productivity Press, 1989, pp. xi, 546, $59.95).

Japanese management is a popular subject in the classrooms of U.S. business schools and in the offices of U.S. corporations. Articles and books that support discussion of the topic abound. Literature on Japanese management accounting practice is not so readily available, however. This book sheds light on how management accounting systems in Japan facilitate managerial decision making in that country.

Japanese Management Accounting is a collection of 33 articles organized in five sections: (1) Cost Management Systems in the Japanese Manufacturing Environment; (2) Profit Planning and Control Systems in Japan; (3) Japanese Cost Accounting Practices and Standards; (4) The Organizational Aspect of Management Accounting in Japan; and (5) Other Topics. The papers are mostly descriptive in nature and include reports of experiences at individual Japanese companies (e.g., Daihatsu Kogyo, Kyocera Corporation, Matsushita, Nippon Steel, and Toyota) as well as generalizations about Japanese management accounting practices based on the results of survey research.

To avoid disappointment, the potential reader of *Japanese Management Accounting* should understand first and foremost what the collection does *not* offer. Such an understanding allows a greater appreciation for what *is* offered. What the book does not do is provide a prescription for applying Japanese management accounting techniques to corporations elsewhere in the world. The articles do not serve as recipes for success. What this collection does do is offer insight. The articles address an exceptionally wide range of issues, including product costing, performance evaluation, transfer pricing, incentive systems, corporate cost allocation, and internal auditing. Taken as a group, the articles show that there is tremendous diversity of practice among Japanese corporations.

Many interesting messages are contained in *Japanese Management Accounting.* One is the extent to which Japanese management accountants have borrowed from their U.S. counterparts. The degree of borrowing is evident from direct acknowledgments, from references to American cost accounting textbooks as justification for practice, and from American textbook-like illustrations of Japanese accounting systems. Not surprising, given the borrowing, is that Japanese management accountants face many of the same problems as those faced by American management accountants. For example, the direct versus absorption costing controversy is alive and well in Japan. Likewise, shifting cost structures have challenged the relevance of labor-based cost systems in Japan, and standard cost systems are playing a lesser role in strategic cost management.

While the collection conveys many similarities between U.S. and Japanese management accounting practices, it also highlights some very real differences. For instance, Japanese companies place much more emphasis on cost reduction in the design phase of the product. A new product's *target cost* (the difference between the expected selling price and the required profit) is established in the initial planning phase. Value engineering activities are applied throughout the preproduction process to ensure that the target cost can be met. If these efforts fail, the product is not introduced because Japanese managers recognize there is little potential for major cost reductions once production begins.

Japanese Management Accounting is worthwhile reading for all management accountants. The numerous surveys comparing Japanese and U.S. practices, while perhaps not so methodologically sound, contain some very interesting statistics. Those surveys and certain of the case analyses, for example, Hamada and Monden's comparison of Kyocera Corporation's amoeba system to the Toyota production system (pp. 208–9), show that there is not complete agreement among Japanese managers concerning how best to achieve maximum efficiency and cost reduction. Several articles in the collection remind the reader of Japan's economic history since World War II. The significant role of the 1973 oil crisis in motivating Japanese managers to adopt just-in-time production systems and other cost reduction activities is not widely publicized in the popular business press, yet is discussed in a number of different contexts throughout *Japanese Management Accounting.*

I recommend this book as supplemental reading at the advanced undergraduate or graduate levels. The articles will undoubtedly generate many interesting classroom discussions. For the professor wishing to internationalize the management accounting curriculum, the collection does much to instill an awareness of the global marketplace; it also reinforces the importance of good management accounting systems for corporations all over the world.

KAY M. POSTON
Assistant Professor of Accounting
Arizona State University–West Campus

WILLIAM V. RUCH, *International Handbook of Corporate Communication* (Jefferson, NC: McFarland & Company, Inc., pp. x, 486, $49.95).

In this age of relative cultural complexity in many countries and of varied methods of communication, any effort to compile a national handbook of corporate communication would be difficult. Mr. William V. Ruch has ambitiously attempted an *International Handbook of Corporate Communication.* He readily admits that the effort which includes the entire nonsocialist world is a mammoth undertaking. I believe the author has made a useful attempt to focus on corporate communication techniques and strategies. He provides the reader significant information on a country-by-country basis.

The author's effort to prepare an *International Handbook of Corporate Communication* is very timely. A perspective of the international economic developments of recent years resulting in a vastly extended international trading network indicates a justification for such a *Handbook.* Although the socialist world is not included, a future handbook would probably include an important section on business communication techniques whether for public or private sector corporations.

The author provides a useful initial chapter, "Communication in the Organization," to indicate the theoretical scope of this major effort. The bulk of the work is devoted to individual country description and analysis with logical grouping of major geographical areas. Specific information on the nature of each country including the land and people, a brief history of the country, its economy, and the organized communication techniques employed in that country are given. Of major importance is the latter section. It includes specific information as to the type and appropriateness of dress for business and for casual appointments, as well as methods in which business communication is conducted. It also provides a useful list of communication techniques that are not utilized or desired in that particular nation.

In many cases the author also includes a case study of a corporation in the particular country being analyzed. The degree of detail varies by country as might be expected. For example, the section of the *Handbook* describing business communication practices in the United Kingdom totals 16 pages of well-detailed notes and illustrative case studies. A smaller and less developed country, such as Lesotho or Botswana, receives a total entry of only one half page.

Important information concerning the specific nation, its business structure, and customs are concisely given and would seem particularly useful to business travelers planning to visit that country for business purposes. This would seem to be the primary public to which the author directs his efforts. There is, I think, for accountants, business students, and practitioners, much useful and interesting information in this work, such as the identification of the various methods of communication, both verbal and nonverbal. The *Handbook* and its approach to corporate communications may also afford some guidance to accountants concerned with effective communication of accounting data.

The work is faithful to its purpose—to serve as a comprehensive reference source for international corporate communication practices. It can serve as a useful guide for the individual seeking information concerning existing culturally effective communication practice in a given country or region. For those seeking a more general concept of the generally accepted business communication practices of various nations, the *Handbook* has three chapters of value. The initial chapter discusses business communication concepts and practices applicable to the Information Age. The final two chapters report the results of the author's international corporate internal communications survey and the general conclusions of the author. The following statement by the author summarizes his conclusion: "... corporate communication programs around the world are remarkably similar in the channels that management chooses to communicate both with management and with one another. Whatever differences do exist can be traced to culture" (p. 445). The book is available from the publisher at P.O. Box 611, Jefferson, NC 28640.

V. K. ZIMMERMAN
Professor of Accountancy

University of Illinois

LARS A. SAMUELSON, *Models of Accounting Information Systems: The Swedish Case* (Lund, Sweden: Student Litteratur, 1990, pp. 220, Skr 160).

This book represents Professor Samuelson's perceptions of the AIS (Accounting Information Systems) experience in Sweden after WWII. His aim is to summarize and interpret the contributions made by Sweden during this time period to the field of AIS. The intended audience for the manuscript is accounting academics and students in doctoral level courses in management accounting. The book is basically divided into three sections and the remainder of this review presents highlights from these sections.

The first section of the book includes two chapters and provides the reader with an understanding of the author's orientation and perceptions of the relevant concepts, terms, and events. In this section he makes a distinction between research *in* accounting and research *of* accounting. This distinction becomes important in the conclusion when Professor Samuelson discusses the book's contributions.

The third chapter begins a section of the book in which the author discusses AIS from different approaches or development models. Each chapter looks at an approach and then presents various historical points at which changes have been made and presents cases to illustrate points made. In some instances the manuscript presented material that is probably not very relevant to most academicians. For instance, in the chapter on the chart of accounts approach, the author spends a few pages discussing the numbers assigned to different accounts under various plans. This same criticism could be made about the next chapter which presents different software packages. He does point out an important change as developers began asking what software should be used as opposed to what chart of accounts is appropriate, but most readers probably will not be interested in the description of the packages that were available.

In contrast, the last chapters in this section present an interesting perspective on the way in which a methodology was formulated to develop AIS. The development of the Swedish RP (or accounting plan) is the result of different groups providing insights and requirements for the methodology. Thus the results may not be without some deficiencies but the author has provided the reader with information about what the results of this process were. The reader comes away from this discussion with an appreciation of the groups that have a stake in the definition of accounting and the various alternatives that were proposed. This perspective provides the logical basis for the section's final chapter in which a different approach to the development of an AIS is called for; one that emphasizes the users' requirements as opposed to methodologies which tend to be based on a top-down centralized approach.

The final section synthesizes the previous chapters and provides the reader with Professor Samuelson's interpretations. In this section the author gives the reader a short overview of what he considers to be the important contributions of each of the topics discussed in the previous chapters and how these events are related to research of and in accounting. Further he presents his view of how the activities in one historical period influenced developments in the next. For instance, he points out that software packages using the V-plan chart of accounts approach were the most successful (p. 178). I believe that this is the type of material that an academic is looking for when they read a book about "The Swedish Case," and I would have preferred that the author had devoted more time to the relationship of this work to other literature and situations. For instance, readers from the U.S. are faced with an interesting difference to consider when the author mentions the impact that various groups have on accounting practice in Sweden which does not have a central body like the SEC to influence (guide) the discipline. Unfortunately, this discussion is very brief.

Initially I assumed that the manuscript would describe the different approaches used to conceptualize and model information systems. When I saw that the author's background was management accounting I expected something else. What I actually found was a book that did not fit into either of these categories, but was a historical description of Sweden's experience in adapting accounting systems to meet their particular needs. It is hard for me to recommend the book to the intended audience because it is not the type of material typically covered in management accounting seminars. In addition, some discussions, such as the discussion of the numbering schemes for different chart of accounts seems extraneous. However, there were certain parts of the book that I found quite interesting and worthwhile. I would have appreciated greater depth in these areas.

GRAHAM GAL
Assistant Professor of Accounting
and Information Systems

University of Massachusetts

GEORGE STALK, JR., and THOMAS M. HOUT, *Competing Against Time—How Time-Based Competition is Reshaping Global Markets* (New York: The Free Press, 1990, pp. x, 285, $24.95).

In this book, two vice presidents of the Boston Consulting Group, the birthplace of portfolio management strategy, detail another strategy, time-based competition, and make a case for it as "the emerging new source of competitive advantage" (p. 28).

The book begins with an overview of the strategies that have dominated the competitive environment since World War II. There are a couple of tantalizing bits in this review from an accounting perspective. The authors cite the mid-1970s as the time when management's attention focused on the relationship between costs and manufacturing, yielding "... a more refined understanding of how costs are driven by varying levels of complexity in an organization..." (pp. 21–2). This claim is not detailed with supporting evidence and it stands in direct contrast to the findings of field research conducted by accounting researchers *even in highly innovative companies* as late as the mid-1980s (e.g., Kaplan 1986, Johnson 1989). In fact, both the section on "de-averaging costs" and "restructuring for advantage" say less about how strategic advantage *was* achieved than how it *could have been* achieved.

The authors use comparative data from U.S. and Japanese companies to illustrate several advantages that accrue to "time compression" of activities. It is interesting to note that, although the claimed focus is on global competition, Europeans seldom enter into the discussion. This is probably because many successful European firms have relied on a strategy other than time-based competition to secure competitive advantage. In the auto industry, one of the authors' favorite sources of examples, the most successful European manufacturers actually take longer than either U.S. or Japanese firms to get new models and major new innovations into their product lines.

The authors devote the middle of the book to fleshing-out the notion of a time-based competitive strategy. They deal with the more well-known supporting activities (elimination of wasted time in the value delivery systems), but they focus their attention, appropriately, on the advantage achievable through flexible and responsive value delivery systems.

The cornerstone of the argument for time-based competition is that there exists a significant number of "demanding customers" (p. 83). This premise is never actually well documented, but the case that such customers will be loyal to a supplier who is responsive to their needs is well argued.

About half of the book is dedicated to the topic of implementing a time-based strategy. The discussions of organizational restructuring and "psychological restructuring" (my term) are by no means exhaustive, and they will be familiar to readers keeping track of the recent trends in manufacturing like just-in-time and total quality control. The discussion includes the typical mechanical tricks for assisting change (benchmarking, new metrics, cross-department mapping) and standard organizational intervention techniques (establishing a vision, identifying barriers to change, using focus groups and pilot programs).

The most useful new information related to implementation is the discussion of the process of gaining competitive advantage by becoming integrated in the value chain. Stalk and Hout examine how an organization can extend its time advantage to help both customers and suppliers become better, more responsive competitors, and hence more loyal. This discussion is essentially about how to use a company's time-based knowledge to best advantage.

This book contains nothing that is particularly new and exciting to a reader who has been keeping pace with the recent literature on manufacturing and strategy. However, the book does do a good job of collecting much of the relevant information and arguments for time-based competition in one place. This makes it particularly useful for readers who are interested in "getting up to speed" on the topics in this literature quickly. Furthermore, the intended audience is obviously business executives, so the style and format is quite easily digestible.

ALFRED J. NANNI, JR.
Associate Professor of Accounting
Boston University

References

Johnson, H. T. 1989. Performance measurement for competitive excellence. Paper presented at the "Measuring Manufacturing Performance Colloquium," Harvard University. (January).

Kaplan, R. S. 1986. Accounting lag: The obsolescence of cost accounting systems. *California Management Review*. (Winter): 174–99.

ROBERT J. THIERAUF, *Expert Systems in Finance and Accounting* (New York: Quorum Books, 1990, pp. x, 277, $49.95).

A typical rule in an expert system might be as follows: If this book *Expert Systems in Finance and Accounting* meets your needs, then buy it. The rule is met if you want or need an intense dose of the more technical aspects of expert systems. Expert systems contain the rules and/or other information regarding what an expert or a group of experts have to say about a particular problem. This book review in many respects is an expert judgment.

There are two primary deficiencies with the book. One, there is no mechanism to follow-up on specific expert systems or even elements of expert systems because few citations are given for the reader. Specifically, in most instances when an example of a system is given, no further information may be found on it. Sometimes there is not even an original citation; thus, follow-up is not available. Second, most of the material in the book is generic to expert systems as opposed to being specific to expert systems for finance and accounting as the title indicates. It appears that the book was written as an introduction to expert systems with expert systems for finance and accounting added on.

Selected elements of the text that are valuable include the introduction to expert systems. It discusses some basic issues and points out the limitations of expert systems. Of lesser value is the amount of time and space devoted to programming languages and expert systems shells.

For readers interested in expert systems programming languages, expert systems shells, and other such technical items, this book offers an overview of those materials. For the reader who is more interested in identifying actual applications and possible uses of expert systems in finance and accounting this book will probably not provide a sufficiently comprehensive coverage nor a sufficient set of working examples to satisfy them.

If the reader desires a demonstration of an expert systems prototype including the knowledge acquisition phase through the demonstration phase, this text leaves something to be desired. A demonstration prototype is offered but it is not easy to follow. Similarly, figures included in the chapter on Operational Prototype and Finance are confusing and seem redundant.

Overall this book offers a broad introduction to expert systems in finance and accounting. It contains significant technical background material but a relatively minimal amount of expert systems applications.

TERRY L. CAMPBELL
Director, Center for Interdisciplinary
Research in Information Systems

Penn State University

BOOKS RECEIVED APRIL 26, 1990 TO AUGUST 8, 1990

CANADIAN INSTITUTE OF CHARTERED ACCOUNTANTS, *Accounting and Reporting for Physical Assets by Governments* (Toronto: CICA, 1990, pp. xiii, 114, CAD $25.00).

CANADIAN INSTITUTE OF CHARTERED ACCOUNTANTS, *The Independent Auditor's Consideration of the Work of Internal Auditors* (Toronto: CICA, 1990, pp. vii, 66, CAD $15.00).

PETER D. CHANT, *The Financial Statement Presentation of Corporate Financing Activities* (Toronto: The Canadian Institute of Chartered Accountants, 1990, pp. xix, 228, CAD $32.50).

FRANK A. COWELL, *Cheating the Government: The Economics of Evasion* (Boston: MIT Press, 1990, pp. xii, 267, $19.95).

DAVID FLINT, *Philosophy and Principles of Auditing: An Introduction* (Houndmills, UK: Macmillan Education Ltd., 1988, pp. xvi, 191, £37.50 hardcover, £11.99 paper).

PETER B. FRANK, MICHAEL J. WAGNER, and ROMAN L. WEIL, Editors, *Litigation Services Handbook: The Role of the Accountant as Expert Witness* (New York: John Wiley & Sons, 1990, pp. xx, 704, $95.00).

M. SHAMSUL HAQUE, *Understanding and Using Lotus Symphony* (Silver Spring, MD: Namuk International, Inc., 1990, pp. xxx, 400, $23.95, paper).

PATRICIA E. MOODY, Editor, *Strategic Manufacturing: Dynamic New Directions for the 1990's* (Homewood, IL: Dow-Jones Irwin, 1989, pp. x, 390, $39.95).

ROBERT J. THIERAUF, *Electronic Interchange in Finance and Accounting* (Westport, CT: Quorum Books, 1990, pp. xiv, 240, $49.95).

UNITED NATIONS CENTER ON TRANSNATIONAL CORPORATIONS, *Advisory Studies Series B No. 6: Curricula for Accounting Education for East-West Joint Ventures in Centrally Planned Economies* (New York: United Nations, 1990, pp. vi, 86, $10.00, paper).

ALBERT S. WILLIAMS, *On Your Own! How to Start Your Own CPA Firm* (New York: American Institute of Certified Public Accountants, 1990, pp. xviii, 213, $35.00, paper).

IRWIN WINSTEN, *Accounting Software and the Microcomputer* (New York: John Wiley & Sons, Inc., 1990, pp. xi, 356, $85.00).

MANUSCRIPT PREPARATION AND STYLE

The *Accounting Review's* manuscript preparation guidelines follows (with a slight modification) the B-format of the *Chicago Manual of Style* (13th ed.; University of Chicago Press). Another helpful guide to usage and style is *The Elements of Style*, by William Strunk, Jr., and E. B. White (Macmillan). Spelling follows *Webster's International Dictionary.*

FORMAT

1. All manuscripts should be typed on one side of 8½ × 11" good quality paper and be double spaced, except for indented quotations.
2. Manuscripts should be as concise as the subject and research method permit, generally not to exceed 7,000 words.
3. Margins of at least one inch from top, bottom, and sides should facilitate editing and duplication.
4. To assure anonymous review, authors should not identify themselves directly or indirectly in their papers. Single authors should not use the editorial "we."
5. A cover page should show the title of the paper, the author's name, title, and affiliation, any acknowledgements, and a footnote indicating whether the author would be willing to share the data (see last paragraph in this statement).

Pagination: *All* pages, including tables, appendices, and references, should be serially numbered. The first section of the paper should be untitled and unnumbered. Major sections may be numbered in roman numerals. Subsections should not be numbered.

Numbers: Spell out numbers from one to ten, except when used in tables and lists, and when used with mathematical, statistical, scientific, or technical units and quantities, such as distances, weights, and measures. For example: *three days; 3 kilometers; 30 years.* All other numbers are expressed numerically. Generally when using approximate terms spell out the number, for example, *approximately thirty years.*

Percentages and Decimal Fractions: In nontechnical copy use the word *percent* in the text; in technical copy the symbol % is used. (See the *Chicago Manual* for discussion of these usages.)

Hyphens: Use a hyphen to join unit modifiers or to clarify usage. For example: a *well-presented analysis; re-form.* See *Webster's* for correct usage.

Key Words: The synopsis is to be followed by four *key words* that will assist in indexing the paper.

SYNOPSIS

A synopsis of about 400 words should be presented on a separate page immediately preceding the text. The synopsis should be nonmathematical and include a readable summary of the research question, method, and the significance of the findings and contribution. The style should be objective, without personal pronouns. The title, but not the author's name or other identification designations, should appear on the synopsis page and on the first page of the text.

TABLES AND FIGURES

The author should note the following general requirements:

1. Each table and figure (graphic) should appear on a separate page and should be placed at the end of the text. Each should bear an arabic number and a complete title indicating the exact contents of the table or figure.
2. A reference to each graphic should be made in the text.
3. The author should indicate by marginal notation where each graphic should be inserted in the text.
4. Graphics should be reasonably interpreted without reference to the text.
5. Source lines and notes should be included as necessary.

Equations: Equations should be numbered in parentheses flush with the right-hand margin.

DOCUMENTATION

Citations: Work cited should use the "author-date system" keyed to a list of works in the reference list (see below). Authors should make an effort to include the relevant page numbers in the cited works.

1. In the text, works are cited as follows: authors' last name and date, without comma, in parenthesis: for example, (Jones 1987); with two authors: (Jones and Freeman 1973); with more than two: (Jones et al. 1985); with more than one source cited together (Jones 1987; Freeman 1986); with two or more works by one author: (Jones 1985, 1987).
2. Unless confusion would result, do not use "p." or "pp." before page numbers: for example, (Jones 1987, 115).
3. When the reference list contains more than one work of an author, the suffix *a*, *b*, etc. follows the date in the text citation: for example, (Jones 1987a) or (Jones 1987a; Freeman 1985b).

4. If author's name is mentioned in the text, it needs not be repeated in the citation; for example, "Jones (1987, 115) says. . . ."
5. Citations to institutional works should use acronyms or short titles where practicable; for example, (AAA ASOBAT 1966); (AICPA Cohen Commission Report 1977). Where brief, the full title of an institutional work might be shown in a citation: for example, (ICAEW *The Corporate Report* 1975).
6. If the manuscript refers to statutes, legal treatises, or court cases, citations acceptable in law reviews should be used.

Reference List: Every manuscript must include a list of references containing only those works cited. Each entry should contain all data necessary for unambiguous identification. With the author-date system, use the following format recommended by the *Chicago Manual*:

1. Arrange citations in alphabetical order according to surname of the first author or the name of the institution responsible for the citation.
2. Use author's initials instead of proper names.
3. Dates of publication should be placed immediately after author's name.
4. Titles of journals should not be abbreviated.
5. Multiple works by the same author(s) should be listed in chronological order of publication. Two or more works by the same author(s) in the same year are distinguished by letters after the date.
6. Inclusive page numbers are treated as recommended in *Chicago Manual* section 8.67.

Sample entries are as follows:

American Accounting Association, Committee on Concepts and Standards for External Financial Reports. 1977. *Statement on Accounting Theory and Theory Acceptance*. Sarasota, FL: AAA.

Fabozzi, F., and I. Pollack, eds. 1987. *The Handbook of Fixed Income Securities*. 2d ed. Homewood, IL: Dow Jones-Irwin.

Demski, J. S., and D. E. M. Sappington. 1989. Hierarchical structure and responsibility accounting. *Journal of Accounting Research* 27: 40–58.

Dye, R., B. Balachandran, and R. Magee. 1989. Contingent fees for audit firms. Working paper, Northwestern University, Evanston, IL.

Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, United Kingdom: Cambridge University Press.

Porcano, T. M. 1984a. Distributive justice and tax policy. *The Accounting Review* 59 (October): 619–36.

———. 1984b. The perceived effects of tax policy on corporate investment intentions. *The Journal of the American Taxation Association* (Fall): 7–19.

Shaw, W. H. 1985. Empirical evidence on the market impact of the safe harbor leasing law. Ph.D. dissertation, University of Texas, Austin.

Sherman, T. M., ed. 1984. *Conceptual Framework for Financial Accounting*. Cambridge, MA: Harvard Business School.

Footnotes: Footnotes are not used for documentation. Textual footnotes should be used only for extensions and useful excursions of information that if included in the body of the text might disrupt its continuity. Footnotes should be double spaced and numbered consecutively throughout the manuscript with superscript Arabic numerals. Footnotes are placed at the end of the text.

SUBMISSION OF MANUSCRIPTS

Authors should note the following guidelines for submitting manuscripts:

1. Manuscripts currently under consideration by another journal or other publisher should not be submitted. The author must state that the work is not submitted or published elsewhere.
2. In the case of manuscripts reporting on field surveys or experiments, four copies of the instrument (questionnaire, case, interview plan, or the like) should be submitted.
3. Four copies should be submitted together with a check in U.S. funds for $50.00 for members or $100.00 for nonmembers of the AAA made payable to the American Accounting Association. Effective January 1990, the submission fee is nonrefundable.
4. The author should retain a copy of the paper.
5. Revisions must be submitted within 12 months from request, otherwise they will be considered new submissions.

COMMENTS

Comments on articles previously published in *The Accounting Review* will be reviewed (anonymously) by two reviewers in sequence. The first reviewer will be the author of the original article being subjected to critique. If substance permits, a suitably revised comment will be sent to a second reviewer to determine its publishability in *The Accounting Review*. If a comment is accepted for publication, the original author will be invited to reply. All other editorial requirements, as enumerated above, also apply to proposed comments.

POLICY ON REPRODUCTION

An objective of *The Accounting Review* is to promote the wide dissemination of the results of systematic scholarly inquiries into the broad field of accounting.

Permission is hereby granted to reproduce any of the contents of the *Review* for use in courses of instruction, as long as the source and American Accounting Association copyright are indicated in any such reproductions.

Written application must be made to the Editor for permission to reproduce any of the contents of the *Review* for use in other than courses of instruction—e.g., inclusion in books of readings or in any other publications intended for general distribution. In consideration for the grant of permission by the *Review* in such instances, the applicant must notify the author(s) in writing of the intended use to be made of each reproduction. Normally, the *Review* will not assess a charge for the waiver of copyright.

Except where otherwise noted in articles, the copyright interest has been transferred to the American Accounting Association. Where the author(s) has (have) not transferred the copyright to the Association, applicants must seek permission to reproduce (for all purposes) directly from the author(s).

POLICY ON DATA AVAILABILITY

The following policy has been adopted by the Executive Committee in its April 1989 meeting.

"An objective of (*The Accounting Review, Accounting Horizons, Issues in Accounting Education*) is to provide the widest possible dissemination of knowledge based on systematic scholarly inquiries into accounting as a field of professional, research, and educational activity. As part of this process, authors are encouraged to make their data available for use by others in extending or replicating results reported in their articles. Authors of articles which report data dependent results should footnote the status of data availability and, when pertinent, this should be accompanied by information on how the data may be obtained."

TATEMENT OF OWNERSHIP, MAN- GEMENT AND CIRCULATION (Re- uired by 39 U.S.C. 3685)

1. *Title of Publication:* The Accounting Re- iew.

2. *Date of Filing:* September 24, 1990.

3. *Frequency of Issue:* Quarterly.

A. *No. of Issues published annually:* 4.

B. *Annual Subscription Price:* $50.00.

4. *Location of known office of publication* treet, city, county, state and ZIP code) (not rinters): 5717 Bessie Drive, Sarasota, Florida 4233-2399.

5. *Location of the headquarters of general* usiness offices of the publishers (not printers): 717 Bessie Drive, Sarasota, Florida 34233- 399.

6. *Names and complete addresses of* ublisher, editor and managing editor.

Publisher (name and address): American Ac- ounting Association, 5717 Bessie Drive, arasota, Florida 34233-2399.

Editor: A. Rashad Abdel-khalik, Fisher chool of Accounting, College of Business Ad- inistration, University of Florida, Gainesville, L 32611.

Managing Editor (name and address): Paul erhardt, 5717 Bessie Drive, Sarasota, Florida 4233-2399.

7. *Owner (If owned by a corporation, its ame and address must be stated and also mmediately thereunder the names and ad- dresses of stockholders owning or holding 1 percent of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)*

Name: American Accounting Association (A nonprofit educational association—no stockholders or "owners").

Address: 5717 Bessie Drive, Sarasota, Florida 34233-2399.

8. *Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities (If there are none, so state):* none.

9. *For completion by nonprofit organizations authorized to mail at special rates (Section 423.12 DMM only). The purpose, function and nonprofit status of this organization and the exempt status for federal income tax purposes:* Have not changed during preceding 12 months.

10. *Extent and nature of circulation:*

A. *Total no. copies printed (Net Press Run).* Average number copies each issue during preceding 12 months: 16,000. Actual number copies of single issue published nearest to filing date: 16,000.

B. *Paid Circulation:*

1. Sales through dealers and carriers, street vendors and counter sales. Average number copies each issue during preceding 12 months: none. Actual number copies of single issue published nearest to filing date: none.

2. *Mail subscriptions.* Average number copies each issue during preceding 12 months: 13,154. Actual number copies of single issue published nearest to filing date: 13,475.

C. *Total paid circulation (sum of 10B1 and 10B2).* Average number copies each during preceding 12 months: 13,154. number copies of single issue published to filing date: 13,475.

D. *Free distribution by mail, carrier means, samples, complimentary, and ot copies.* Average number copies each during preceding 12 months: 37. Actual copies of single issue published nearest t date: 37.

E. *Total distribution (Sum of C* Average number copies each issue duri ceding 12 months: 13,191. Actual copies of single issue published nearest t date: 13,512.

F. *Copies not distributed:*

1. *Office use, left over, unaccounted, after printing.* Average number copie issue during preceding 12 months: 2,809. number copies of single issue published to filing date: 2,488.

2. *Returns from news agents.* number copies each issue during prece months: none. Actual number copies o issue published nearest to filing date: n

G. *Total (Sum of E, F1 and 2—shou net press run shown in A).* Average copies each issue during preceding 12 16,000. Actual number copies of sing published nearest to filing date: 16,000.

H. *I certify that the statements mad above are correct and complete.*

Paul L. Gerhardt, Executive Direct

CONTENTS OF ACCOUNTING RESEARCH JOURNALS

Accounting and Business Research

A research quarterly published by the Institute of Chartered Accountants in England and Wales

Number 80 Autumn 1990

Editors: R. H. Parker, University of Exeter
C. W. Nobes, University of Reading

CONTENTS

Multinational Performance Measures and Their Association with Contextual Variables	Istemi S. Demirag
Some Issues in Corporate Hedging Policy	S. Eckl J. N. Robinson
Standard Setting for the Financial Reporting of Religious Business Organisations: The Case of Islamic Banks	Rifaat Ahmed Abdel Karim
The Accounting Burdens Facing Small Firms: An Empirical Research Note	Kevin Keasey Helen Short
A Longitudinal Study of Accounting Changes: The UK Gas Industry 1969–1974	W. M. McInnes
Accounting Standards and the Judiciary	G. W. Radcliffe
Origins of the Plan Comptable General: A Study in Cultural Intrusion and Reaction	Peter E. M. Standish
The Use of Computers in Accounting Course: A New Perspective—A Comment	P. A. Collier G. R. Kaye B. J. Spaul B. C. Williams
Restricting the Domain and Potential of Cash Flow Accounting	T. A. Lee
Book Reviews	

Subscriptions should be sent to 40 Bernard Street, London WC1N 1LD, England.

JOURNAL OF ACCOUNTING & ECONOMICS

Table of Contents

Volume 13, Number 3 September 1990

Editorial Data

Douglas J. Skinner
Options Markets and the Information Content of Accounting Earnings Releases

Daniel W. Collins and Linda DeAngelo
Accounting Information and Corporate Governance: Market and Analyst Reactions to Earnings of Firms Engaged in Proxy Contests

Shehzad L. Mian and Clifford W. Smith, Jr.
Incentives Associated with Changes in Consolidated Reporting Requirements

John S. Strong
Valuation Effects of Holding Gains on Long Term Debt

Brett Trueman
Theories of Earnings Announcement Timing

Advertisements

JOURNAL OF ACCOUNTING AND PUBLIC POLICY

VOLUME 9 NUMBER 2 SUMMER 1990

Contents

Financial Disclosure Regulation and Its Environment: A Review and Further Analysis 79
Terence E. Cooke and R. S. Olusegun Wallace

An Empirical Analysis of Management's Choice of Accounting Treatment for Business Combinations 111
Kathleen M. Dunne

Privileged Communications, Independence, and Client Advocacy: A Perspective on Professionalism 135
G. Stevenson Smith

Biographies 159

Journal of Accounting and Public Policy is published quarterly by Elsevier Science Publishing Co., Inc., 655 Avenue of the Americas, New York, NY 10010.

All editorial correspondence should be sent to: Lawrence A. Gordon or Stephen E. Loeb, College of Business and Management, University of Maryland at College Park, College Park, MD 20742.

parison with the empirical results on error causes (in manufacturing) reported in Libby (1985) and Libby and Frederick (1990). In both studies, the relation between error frequency knowledge and several measures of experience is examined.

This article provides evidence on three research questions: (1) *How many audits (in a particular industry) does an auditor experience?* The answer to this question will help establish whether auditors learn error frequencies from direct, personal experience with financial statement errors or acquire their knowledge of error frequencies by other means. (2) *What do auditors know about the relative frequencies actually associated with the population of financial statement errors discovered during the audit process?* If error frequency knowledge is essential to audit expertise, then it is useful to understand the nature of that knowledge. Further, if expertise is to be measured, a valid empirical measure of knowledge must be developed. (3) *Do more experienced auditors have more accurate error frequency knowledge than less experienced auditors?* Discernible knowledge differences across experience levels would suggest that error frequency knowledge is gained through audit experience, consistent with the general psychological characterization of expertise.

The results show, first, that even the most experienced auditors have limited direct experience with financial statement errors. Second, auditors seem to know only the *most frequently occurring* error effects and causes. Third, differences in auditors' knowledge of error effects across experience levels are not explained by differences in the length of either audit experience or industry-specific audit experience, or by the number of clients audited in an industry. Moreover, auditors with similar experience levels show large individual differences in knowledge of causes and effects.

These results suggest that audit experience should be viewed as relating to specific audit tasks rather than as a singular, all-encompassing concept and that particular experience must be understood as it relates to a particular type of knowledge. Moreover, the results and the fact that financial statement errors are rare events raise questions about the value of normative and descriptive models that have been proposed as the basis for understanding audit judgment and expertise—models that assume knowledge of error frequencies.

Key Words: *Expertise, Experience, Knowledge, Error frequency.*

Data Availability: *Inquiries about the data collection materials and data availability should be addressed to the author.*

THE psychological literature on expertise offers two general conclusions. First, domain-specific knowledge is the essential determinant of expertise. Second, expert knowledge is gained through many years of on-the-job experience (see, e.g., Chi et al. 1983). Auditing research has implicitly or explicitly adopted these conclusions as fundamental assumptions in attempts to establish a knowledge/experience/expertise link in auditing. In addition, auditing literature suggests that *knowledge about error frequencies* is essential domain-specific knowledge in auditing. Previous research has supported the hypothesis that error causes that are frequently detected in practice are perceived by auditors to occur more frequently (Libby 1985, 651) and the additional hypothesis that as experience increases, perceptions of the frequency of financial statement error causes become more accurate (Libby and Frederick 1990).

This paper reconsiders and extends these hypotheses. It is organized as follows. The first section describes literature relevant to understanding the assumptions on which this research is based, while the second section describes the designs and results of the two studies. The paper concludes with a general discussion of the results and directions for further research.

I. Relevant Literature

Experience Required for Expertise

The nature and amount of experience required for expertise in auditing is in question.[1] A striking discrepancy exists between the amount of direct audit experience possessed by "expert" auditors and that of experts in other domains. Researchers have used the number of years in public accounting or the hierarchical level in the firm (i.e., staff through partner) as a surrogate for audit expertise and experience. The time required for this designation has varied from about three to five years.

Research on expertise in other fields suggests that *extensive* experience—through years of intensive practice and repetition—is necessary for expertise. For example, it is estimated that master chess players have spent about 30,000 hours at the game (Simon and Chase 1973). Similarly, radiologists see about 10,000 x-rays during their medical training and about 200,000 in a career (Lesgold et al. 1988, 312); physicians typically see 25 patients a day (Owens 1983, 106).

In contrast, auditors experience relatively few audits—about ten to 15 per year (Abdolmohammadi 1987, 182). In addition, discovered errors in financial statements are rare events. Thus, an important issue is whether and how much to extrapolate from the general expertise literature to auditing. This article provides initial empirical evidence on the nature and amount of auditors' experience.[2]

---

[1] Audit experience is conceptualized as a continuum from *direct* experience (e.g., participating in an audit in which an error is found) to *indirect* experience (e.g., receiving summary data on error frequency). Sources of *intermediate* experience include firm training exercises, informal conversations with colleagues, and firm literature (see Butt 1988).

[2] Important differences between experience in auditing and experience in chess, radiology, and medicine also include the absence of feedback, due to the absence of a criterion for most audit judgment tasks. When there are explicit criteria for success, experts can be reliably differentiated from novices in research studies on bases other than experience or knowledge. For example, master chess players are distinguished from novices based on *demonstrations* of who consistently wins the games. Also consider Einhorn's (1972) study in which expert medical pathologists predicted survival times for patients who suffered from Hodgkin's disease. The cases studied were those in which patients had died, providing an objective measure of the pathologists' expertise in predicting survival times. In auditing, however, the absence of a reliable independent criterion for identifying experts is troublesome. Further, auditing expertise is likely to differ in that knowledge levels are defined by the quality, rather than the amount, of experience. For example, Jacoby et al. (1986, 469) observe that: "While drinking twenty six-packs a day may make one experienced, heavy and drunk, it would not necessarily make one an expert on beer."

Frequency Knowledge as a Component of Audit Expertise

Because of the link between descriptive models of audit judgment and normative decision theory—in which error base rates are revised based on client-specific information in evaluating the likelihood that an error is present in an account—the importance of frequency knowledge to audit expertise has long been *assumed* in the literature. For example, Kinney (1975, 118) considers the likelihood of errors in a single account balance as he writes, "The auditor is uncertain as to what state is or will be the true state, but has some knowledge of what state is likely to occur by virtue of his general and specific knowledge and experience in such matters." From the perspective of the memory and expertise literature, Libby (1983, 5) asserts that "[w]hether auditors want to or not, they acquire and automatically encode information such as the frequencies of occurrence of different financial statement errors they come across."

A specific empirical link between error frequency knowledge and expertise is developed as part of two studies in Libby (1985) and Libby and Frederick (1990). Two hypotheses from those studies are reexamined in the present research. Libby (1985) tested the hypothesis that financial statement error causes that are detected more frequently in practice are perceived to occur more frequently. Auditors ranked the relative frequencies of a set of financial statement error causes, e.g., "bad debt expense and allowance not recorded or underrecorded," or "sales to valid customers are recorded more than once." The results show that auditors' mean judgments of error-cause frequencies are highly correlated (r=0.82) with the actual frequencies from a sample of audit workpapers from 82 Peat Marwick manufacturing clients (Willingham and Wright 1985). Libby concluded that "the mean participant judgment was quite accurate" (1985, 658).[3]

The second hypothesis (Libby and Frederick 1990, 4, 8) suggests that auditors learn error occurrence rates as they gain experience. Accuracy for *each* subject is measured by the correlation between his or her perceived frequencies and the actual frequencies in the task used in Libby (1985). The results indicate that experience is significantly related to accuracy: Audit managers with a median of five years of experience are more accurate than audit staff with a median of one year of experience. The staff auditors are, in turn, more accurate than a sample of student subjects. The mean accuracy levels are about 0.58 for the managers, 0.48 for the staff, and 0.26 for the students. These results suggest that acquisition of knowledge about error cause frequency is the result of experience in auditing, although the accuracy levels for the managers are lower than in Libby (1985).[4]

Thus, the Libby (1985) and Libby and Frederick (1990) articles provide the first empirical results on knowledge of financial statement error causes. The present study reexamines experienced auditors' knowledge of error-cause frequencies at the level of the individual auditor. It reveals specifically which error frequencies they do and do not know. In addition, assuming that if auditors learn error-cause frequencies through experience they also learn error-effect frequencies, the study examines auditors' knowledge of error-effect frequencies across industries and experience levels. It also compares, within subjects, error-effect and error-cause frequency knowledge for the manufacturing industry.

---

[3] Note that this is not "mean accuracy of participants," the traditional measure of accuracy that averages individuals' correlational accuracy scores. Averaging the perceived frequencies increases measured accuracy because of the reduction in random error (e.g., Libby and Blashfield 1978). Libby points out that the standard deviations of these frequency judgments are quite high, and he expresses concern about individual auditor inaccuracy: "If this variation is not due to measurement error or to the idiosyncrasies of each auditor's client mix, it could be indicative of possible decision errors or inefficient strategies. Recent attempts by audit firms to accumulate archival frequency data may provide a useful supplement to the auditor's experience, particularly to those with less relevant experience" (1985, 664).

[4] The accuracy levels in Libby and Frederick (1990) are calculated differently than in Libby (1985) since the former paper averages accuracy scores rather than frequency judgments.

Frequency Knowledge Acquisition through Experience

Research in psychology suggests that acquisition of frequency knowledge is automatic (Hasher and Zacks 1979, 1984). Unlike many other types of knowledge, frequency knowledge is not acquired through voluntary, effortful strategies, nor is it influenced by training, ability, motivation, or intention. Even when individuals are not aware of their frequency knowledge, they have keen ability to discriminate differences in the relative frequencies of events (Hasher and Zacks 1984). In fact, acquisition is so reliably automatic that knowledge about frequency information can be used as a research tool to answer questions about memory and attention (Hock and Hasher 1990).

Evidently auditors attend to frequency information associated with errors experienced during audits. Indeed, information about the frequency with which errors affect financial statements is formally summarized in the audit workpapers. Thus, auditors' knowledge of error frequencies should be a reliable index of their experience with errors. Psychological research that supports these assertions are exemplified by Attneave (1953), who found that subjects' knowledge of the frequency with which different letters of the alphabet occur correlates 0.88 with actual frequency. Similar results (in the 0.70 to 0.90 range of correlation) indicate that individuals are able to accurately estimate the frequency of occurrence of common words, surnames, professions, syllables, letter combinations, and sources of mortality and morbidity. Hasher and Zacks (1984) observe that "these are not the sorts of events whose frequency one might be expected to learn deliberately."

One implication of this research is that if auditors acquire accurate error frequency knowledge through experience, there will be a difference in the accuracy of that knowledge between the more and the less experienced auditors. Specifically, a positive relation should exist between accuracy and experience. On the other hand, the absence of a positive relation between accuracy and experience would suggest two possibilities: (1) To the extent that auditors exhibit highly accurate frequency knowledge, knowledge is acquired vicariously early in the auditors' careers and is unaffected by direct experience with additional audits. (2) Uniformly inaccurate error frequency knowledge across experience levels suggests that auditors have not experienced a representative sample of financial statement errors.

II. The Two Studies

A. Overview

The first study involves 453 Peat Marwick auditors from 26 practice offices who have widely varying levels of experience in five industries. Auditors from all levels in the firm (i.e., partner, senior manager, manager, supervising senior, senior, and staff) assessed error effects—the frequency with which errors affected financial statement accounts—in the five industries.

Participants in the second study are 19 Peat Marwick auditors with an average of more than five years of experience (in supervising senior, manager, and senior manager positions) who are on rotational assignment in the firm's executive offices. The assessment of error effect frequencies is repeated in this study for the manufacturing industry only. The auditors in the second study also assess the relative frequencies associated with *causes* of errors in manufacturing industry accounts.

In both studies, accuracy, or knowledge level, is measured by comparing each auditor's assessed error frequencies with the actual frequencies of discovered errors based on archival data. The accuracy criterion for error effects is obtained from a representative sample of 186 Peat Marwick audits in the manufacturing, merchandising, natural resources, banking, and insurance industries; Wright and Ashton (1989) provide the actual number of errors that affected each balance sheet and income statement account for each of the five industries. Similarly, the criterion for error causes in manufacturing is data from 82 Peat Marwick manufacturing clients collected by Willingham and Wright (1985) and used as the accuracy criterion in Libby (1985) and Libby and Frederick (1990).

B. The First Study

The Accuracy Criterion. The Wright and Ashton (1989) data were provided by audit managers who identified the four largest proposed audit adjustments that are not below 20 percent of "gauge," an explicit materiality measure used by Peat Marwick for audit planning purposes (Elliott 1983). Adjustments of a "bookkeeping" nature that are performed as a client service such as anticipated adjustments to deferred and accrued income taxes were excluded. Responses from 186 engagements provided information on 368 proposed adjustments, or error effects.

Procedure. The data collection instrument was distributed by the firm's executive offices over the signature of John Willingham, Director of Research.[5] Materials were sent to partners-in-charge of audit practice in the 26 largest domestic offices. Large offices were selected to increase the probability that industry specialists with substantial experience levels would be included in the sample, and the geographical distribution of the offices ensured adequate representation of the five industries. Completed materials were returned to the executive office and then forwarded to me.

The Sample of Auditors. Although 499 auditors returned frequency data, the results are based on usable responses by 453 auditors.[6] Of the 447 participants who indicated their level in the firm, there are 48 staff, 81 seniors, 93 supervising seniors, 91 managers, 97 senior managers, and 37 partners.

Data Collection Instrument. The data were elicited by a packet of materials that took 30 to 45 minutes to complete. Three pages of general information described the sample of audits to which the auditors' responses would be compared. Next, a page of specific instructions regarding the error frequency responses was followed by five pages (one for each industry) used to elicit the participant's estimates of error frequencies. For each industry, 100 points were allocated to the financial statement accounts to indicate the auditor's estimate of the percentage of total discovered errors affecting the account categories for that industry.

---

[5] A cover letter described the research, encouraged participation, asked that the instrument be distributed to auditors with a broad range of experience, and stated that auditors who had jointed the firm via the recent merger with Main Hurdman should be excluded from participation. (Since the merger occurred in the interim between the completion of Wright and Ashton's (1989) data collection and that for the present study, the former Main Hurdman employees had only limited experience with Peat Marwick clients.)

[6] Forty-six responses were not analyzed for one of the following reasons: (1) participants failed to comply with the requirements of the task by treating income statement accounts as offsets to balance sheet accounts, thereby failing to assess separate frequencies for the two account types (18 responses); (2) participants had gained the majority of their experience in a firm other than Peat Marwick (18 responses); and (3) participants failed to provide complete information regarding months of audit experience (10 responses).

Accounts affected by extremely small numbers of errors were grouped in a category of "other accounts." The total percentage of errors affecting those accounts was stated and the names of the accounts were provided in the instructions for each industry. The account titles used for the manufacturing, merchandising, and natural resources industries were identical, and participants were required to allocate approximately 95 points (depending on the percentages stated for the "other" accounts) to 14 accounts (ten balance sheet and four income statement) in each of those three industries. Different charts of accounts were required for the banking and insurance industries.[7] Participants allocated 91 (98) points to ten (11) account categories for banking (insurance). The materials concluded with three pages requesting information about the participant's education and auditing experience.

Analysis. Data analysis was done at both the group and the individual auditor levels. Accuracy is measured as the Pearson correlation between actual error frequencies and the estimates of those frequencies by the auditor. The manufacturing industry accuracy score for each auditor is the correlation between the 14 actual error frequencies for that industry and the 14 estimates of those frequencies by the auditor. Accuracy for the other four industries is measured in the same way.

To compare accuracy with experience, the latter is operationalized in two ways. It is measured discretely by level or position in the auditing firm (partner, senior manager, manager, supervising senior, senior, and staff), and continuously by length of service (in months of experience and by number of clients audited). The number of months of industry experience is the auditor's self-reported number for industry-specific auditing experience.

The individual-level analyses provide a finer perspective on the general question of what individual auditors know (and do not know) with respect to relative error frequencies. They show how many auditors ranked each error first, second, and so on. Thus, it becomes clear which errors are ranked correctly by most auditors and what misrankings occur, and how frequently.

Group Results

Industry Accuracy and Industry-Specific Experience. First, I consider the relationship between industry-specific experience and accuracy.[8] The discrete analysis in table

---

[7] Wright and Ashton (1989) report their results using only one set of financial statement accounts—those appropriate for the manufacturing, merchandising, and natural resources industries. The authors provided more detailed data for this study, so that actual frequencies for accounts in the banking and insurance industries could be assessed.

[8] A total of 63 error frequency estimates was required for the five industries (i.e., 14, 14, 14, 10, and 11 for manufacturing, merchandising, natural resources, banking, and insurance, respectively). Based on 344 auditors who responded for all five industries—whether or not they had industry experience—the mean overall accuracy score by experience level was as follows (n=number of auditors): staff 0.34, $n=40$; senior 0.32, $n=60$; supervising senior 0.34, $n=78$; manager 0.34, $n=71$; senior manager 0.33, $n=66$; and partner 0.34, $n=29$. The correlation between accuracy by industry (for all auditors who responded for the industry—again, without regard to experience) and months of Peat Marwick experience (n=number of auditors) is: manufacturing 0.02, $n=434$; merchandising 0.12, $n=404$; natural resources 0.01, $n=366$; banking 0.08, $n=384$; insurance 0.01, $n=364$.

Table 1
First Study: Relation Between Accuracy and Present Position in the Firm for Auditors with Industry-Specific Experience

Position	Number of Auditors	*Accuracy*[a]			
		All Accounts		*Balance Sheet Accounts Only*[b]	
		Mean	*Std. Dev.*	*Mean*	*Std. Dev.*
		Manufacturing			
Staff	38	0.50	0.16	0.75	0.17
Senior	70	0.46	0.19	0.70	0.24
Sup. Sr.	72	0.51	0.18	0.73	0.19
Manager	84	0.50	0.20	0.74	0.21
Sr. Mgr.	89	0.48	0.20	0.74	0.18
Partner	30	0.51	0.16	0.77	0.14
		Merchandising			
Staff	21	0.46	0.24	0.62	0.25
Senior	30	0.45	0.22	0.63	0.23
Sup. Sr.	40	0.49	0.16	0.63	0.15
Manager	50	0.49	0.18	0.66	0.21
Sr. Mgr.	60	0.46	0.14	0.65	0.16
Partner	23	0.52	0.09	0.74	0.11
		Natural Resources			
Staff	6	0.23	0.36	0.18	0.51
Senior	19	0.26	0.29	0.24	0.34
Sup. Sr.	31	0.25	0.24	0.25	0.29
Manager	18	0.22	0.28	0.18	0.35
Sr. Mgr.	29	0.18	0.27	0.15	0.34
Partner	9	0.21	0.28	0.21	0.36
		Banking			
Staff	37	0.26	0.14	0.32	0.15
Senior	50	0.20	0.18	0.27	0.18
Sup. Sr.	57	0.22	0.15	0.31	0.17
Manager	61	0.17	0.19	0.24	0.18
Sr. Mgr.	41	0.24	0.19	0.30	0.19
Partner	23	0.26	0.15	0.34	0.16
		Insurance			
Staff	12	0.17	0.26	0.08	0.33
Senior	28	0.12	0.22	0.13	0.23
Sup. Sr.	27	0.18	0.26	0.16	0.25
Manager	27	0.15	0.24	0.14	0.27
Sr. Mgr.	28	0.15	0.28	0.12	0.31
Partner	18	0.11	0.20	0.05	0.35

[a] Accuracy is measured as the Pearson correlation between an individual's responses and the actual error frequencies for an industry.

[b] I do not present "Income Statement Accounts Only" data because they repeat the data presented here (i.e., they explain the difference, if any, between data for All Accounts and for Balance Sheet Accounts Only).

Table 2
First Study: Relation Between Accuracy[a] and Two Continuous Experience Variables for Auditors with Industry-Specific Experience

	Industry Experience in Months				*Number of Clients*			
		Number of Months				*Number of Clients*		
Industry[b]	*Number of Auditors Responding*	*Mean*	*Std. Dev.*	*Correlation Between Acc., Months*	*Number of Auditors Responding*[c]	*Mean*	*Std. Dev.*	*Correlation Between Acc., #Clients*
Mfg.	383	30	63	0.04	335	9	24	0.02
Mer.	224	15	31	0.04	197	4	7	−0.02
NR	112	19	28	0.27	98	4	5	0.29
Bkg.	269	20	31	0.11	229	6	12	0.10
Ins.	140	17	36	−0.17	125	4	5	−0.21

[a] Accuracy (Acc.) is measured as the Pearson correlation between an individual's responses and the actual error frequencies for an industry.

[b] The industries included are manufacturing, merchandising, natural resources, banking, and insurance.

[c] The number of participants differs because not all participants indicated the number of clients with which they have experience.

1 considers the relationship between industry accuracy and the participant's present position in the firm. Only auditors who assessed error frequencies for that industry and reported at least one month of experience in the industry are included. Table 1 provides the mean accuracy measures for each industry by position in the firm. When all accounts (both balance sheet and income statement) are considered, mean accuracy ranges from 0.11 for Insurance Partners to 0.52 for Merchandising Partners. Within each industry, however, there is no clear pattern of increasing accuracy for individuals with more experience. However, mean accuracy is generally greater for the manufacturing and merchandising industries than for the natural resources, banking, and insurance industries. Table 1 also indicates no systematic pattern in the variability of accuracy across experience levels. Thus, like accuracy, consensus does not appear to be related to experience.

Table 1 also shows that auditors in the manufacturing and merchandising industries generally were more accurate in assessing error frequencies for balance sheet accounts than for all accounts combined (i.e., balance sheet and income statements).[9] Otherwise, no consistent relationship between accuracy and experience as measured by position in the firm emerges from this analysis.

The relationship between industry accuracy and industry-specific experience using the two continuous experience variables (industry experience in months and number of clients audited) is described in table 2. The analysis includes only those auditors who assessed error frequencies for that industry *and* reported at least one month of audit experience in the industry. The number of auditors varies slightly between the two experience variables because some auditors did not report the number of clients they had audited in an industry.

---

[9] Further analyses of these data (presented later) and the results of the second study elaborate this finding.

A FORUM ON
COGNITIVE RESEARCH IN AUDITING

1. Alison Hubbard Ashton

Experience and Error Frequency Knowledge as Potential Determinants of Audit Expertise

(Submitted: August 1988. Accepted: October 1990.)

2. David M. Frederick

Auditors' Representation and Retrieval of Internal Control Knowledge

(Submitted: June 1988. Accepted: October 1990.)

3. Eric E. Spires

Auditors' Evaluations of Test-of-Control Strength

(Submitted: July 1988. Accepted: November 1990)

4. Robin M. Hogarth

A Perspective on Experience and Expertise in Auditing

(Commissioned by the Editor. Final draft received: January 1991)

Editor's *Note:* A "Forum" is a collection of articles around a common theme. Accepted refereed articles are followed by an invited "perspective" authored by experts on that theme. In addition to allowing for synergy and to providing alternative explanations for the results, an objective of the "perspective" is to provide guidance for future research on issues related to that common theme.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 218–239

Experience and Error Frequency Knowledge as Potential Determinants of Audit Expertise

Alison Hubbard Ashton
Duke University

SYNOPSIS: Analyses of audit judgment have suggested or implied that frequency knowledge (i.e., knowledge of the base rates associated with error occurrences) acquired through experience is an important component of audit expertise. For example, auditors are assumed to use base-rate expectations about error occurrence (modified to reflect client-specific information) to allocate audit effort among financial statement accounts and to use likelihoods associated with alternative causes of errors that are identified by analytical review. This article reports two studies that examine both *error-effect* frequency knowledge, i.e., the frequency with which an individual financial statement account is affected by error, and *error-cause* frequency knowledge, i.e., the underlying reason for an error in a particular account.

In the first study, auditors' knowledge of the frequencies with which errors affect financial statement accounts in five industries is compared with archival frequencies. In the second study, auditors' knowledge of both the causes and effects of errors in the manufacturing industry is compared with the archival frequencies of both error causes and their effects on financial statement accounts. The second study provides a within-auditor comparison of both types of error frequency knowledge and allows com-

Support for this research was provided by The Accounting Center Fund in the Fuqua School of Business, Duke University. I am grateful to Arnold Wright and Robert Ashton for supplying their data and for helpful insights in the design of this study, to John Willingham for advice and assistance in distributing the data collection materials, and to Nancy Hubbard for pilot testing the materials. The paper has benefitted from discussions with James Bettman, Elizabeth Davis, William Felix, David Frederick, Lynn Hasher, Vicky Heiman, Robert Libby, Laureen Maines, William Messier, John Payne, David Plumlee, Steven Salterio, Joseph Schultz, Ira Solomon, and William Wright, as well as participants in accounting workshops at Duke University, the University of Florida, and the USC Audit Conference. Research assistance by Michael Guiry, Oris Stuart, and Robert Ching is gratefully acknowledged.

Submitted August 1988.
Revisions received October 1989, August 1990, and October 1990.
Accepted November 1990.

As before, no support for a consistent relationship between accuracy and experience is found. The correlations between accuracy and number of months of industry experience range from −0.17 (insurance) to 0.27 (natural resources). The correlations between accuracy and number of clients are similar, ranging from −0.21 (insurance) to 0.29 (natural resources).

Audit Experience. The experience data reported in table 2 are described more completely in table 3. Except for senior managers and partners in manufacturing and partners in banking, all experience level and industry combinations have an average of nine or fewer clients audited. Also, in most cases only one or two clients separate the average number of clients from rank to rank. Few participants in this study have auditing experience with more than 20 clients in an industry. Thus, the opportunity to observe error events is quite limited for most auditors, even for those who occupy more senior positions in the firm.

Individual Results

The group-level results presented above reveal no relation, on average, between accuracy and audit experience. In this section, the responses of the *individual* auditors is reported as a "confusion matrix" (Attneave 1953, 1959; Pitz 1979) for each industry. The matrix for manufacturing is shown in table 4. The results for other industries are described below.

The rankings of error frequencies for balance sheet and income statement accounts are reported separately as panel A and panel B in table 4. The actual ranks for the errors as determined in Wright and Ashton (1989) are shown across the top of the matrix. The ranks assigned by auditors are shown down the left side of the matrix. The cell entries represent the number of auditors who ranked the error in that cell. Thus, if all auditors had been perfectly accurate in assessing error frequencies for the manufacturing industry, the diagonal cell entries in the balance sheet and income statement matrices in table 4 would be 383 each. Thus, the confusion matrices show the financial statement errors for which individual auditor knowledge is more or less accurate.

Panel A: Balance Sheet Accounts. Because the balance sheet accounts are so numerous, it is difficult to discern the pattern of accuracy versus "confusion" in these matrices. Therefore, at the bottom of the matrices, the "%" row provides a summary measure of the percentage of auditors who are "approximately" accurate in ranking the error frequencies. The percentage is calculated as follows: The observations in the diagonal cell and the cells immediately preceding and following the diagonal cell in each column are summed to represent the responses that are correct or are within one rank (plus or minus) of being correct. (Note that for ranks one and ten, only two cells can be summed.) This sum is divided by the total number of auditors included in the matrix. Thus, the 0.95 in the leftmost column of the balance sheet matrix in table 4 is (318+46)/383, and the 0.86 in the second column is (70+184+74)/383. Tied actual ranks are indicated by equals signs across the top of the matrix and are described in footnote b to the table. The number of cells included in the percentage calculation is expanded for tied ranks. For example, in manufacturing, for which actual ranks five and six are tied, cells four through seven are summed in the numerator of the percentage calculation.[10]

---

[10] The "%" rows for the balance sheet accounts for other industry matrices (omitted due to space constraints) are Merchandising, 0.92, 0.07, 0.67, 0.60, 0.11, 0.54, 0.66, 0.72, 0.78, and 0.42; Natural Resources, 0.47, 0.63, 0.29, 0.05, 0.30, 0.56, 0.48, 0.79, 0.13, and 0.04; Banking, 0.30, 0.22, 0.13, 0.41, 0.61, 0.38, 0.18, and 0.41; and Insurance, 0.94, 0.02, 0.26, 0.41, 0.65, 0.16, and 0.20.

Table 3
First Study: Auditors' Industry Experience by Position

		Number of Clients		Months of Experience	
Position	Number of Auditors[a]	Mean	Std. Dev.	Mean	Std. Dev.
		Manufacturing			
Staff	35	3	3	5	4
Senior	60	5	7	9	6
Sup. Sr.	69	5	4	16	24
Manager	66	6	8	21	18
Sr. Mgr.	77	10	11	44	42
Partner	27	28	32	113	111
		Merchandising			
Staff	17	2	1	3	2
Senior	25	3	3	5	4
Sup. Sr.	39	2	3	8	9
Manager	43	3	4	10	12
Sr. Mgr.	51	6	10	27	40
Partner	23	6	11	46	70
		Natural Resources			
Staff	4	2	1	6	4
Senior	16	2	1	9	11
Sup. Sr.	29	4	3	15	15
Manager	15	4	5	19	20
Sr. Mgr.	25	9	8	26	31
Partner	9	6	6	57	64
		Banking			
Staff	30	3	2	4	3
Senior	46	4	8	8	8
Sup. Sr.	53	3	3	11	10
Manager	50	5	5	18	18
Sr. Mgr.	35	7	7	31	37
Partner	20	14	16	52	57
		Insurance			
Staff	12	1	1	2	1
Senior	24	2	2	6	6
Sup. Sr.	25	2	1	6	8
Manager	26	3	3	12	15
Sr. Mgr.	24	6	6	23	25
Partner	17	9	11	59	78

[a] Numbers are less than or equal to those in table 1 because not all auditors provided both these experience data.

Table 4
First Study: Individuals' Knowledge for the Manufacturing Industry[a]
(n=383)

Panel A. Balance Sheet Accounts:

Assigned Rank	Actual Rank[b] 1	2	3	4	5	=	6	7	8	9	10	Total
1	318	70	4	15	1		1	1	1	1	18	430
2	46	184	46	33	6		2	2	4	7	35	365
3	10	74	96	90	22		7	4	9	22	65	399
4	2	25	64	73	32		9	18	20	28	54	325
5	2	14	54	59	38		15	33	42	43	57	357
6	4	9	41	38	56		20	50	51	54	37	360
7	1	6	38	50	87		75	94	96	94	52	593
8	0	1	31	20	72		91	97	96	92	42	542
9	0	0	7	5	63		87	77	52	42	20	353
10	0	0	2	0	6		76	7	12	0	3	106
%[c]	0.95	0.86	0.54	0.58	0.56		0.31	0.63	0.64	0.35	0.06	3,830

Panel B. Income Statement Accounts:

Assigned Rank	Actual Rank 1	2	3	4	Total
1	13	267	133	9	422
2	66	85	172	58	381
3	227	26	65	237	555
4	77	5	13	79	174
					1,532

---

[a] Indicates the number of auditors who assigned each rank to the account represented by the actual rank. If all auditors had ranked the accounts perfectly accurately, the diagonal cells would contain 383, and all other cells would contain zero.

[b] Actual ranks correspond to the following accounts: *Balance Sheet* 1. Inventory; 2. Accounts receivable; 3. Accounts payable; 4. Accrued liabilities; 5. and 6. Property, plant, and equipment and Stockholders equity have identical frequencies; 7. Other current liabilities; 8. Prepaid expenses and other current assets; 9. Deferred charges and other noncurrent assets; 10. Deferred taxes. *Income Statement* 1. General and administrative expenses; 2. Cost of goods sold; 3. Revenue; 4. Selling expenses.

[c] Indicates the percent of total (n=383) number of auditors who ranked the error in the correct category or in the categories one above or one below the correct category.

The confusion matrices suggest some interesting generalizations about auditor knowledge of error frequencies. For balance sheet accounts, in three of the industries (manufacturing, merchandising, and insurance), over 75 percent of the auditors with some industry experience rank the most frequently affected account first. In addition, more than 90 percent of the auditors classify the most frequently affected account as either the first or second most frequently affected. However, in these three industries fewer than half of the auditors rank the second most frequently affected account second. In the third and later ranks for these industries there is little consistency in the patterns of rankings across industries, except for the general finding that error frequencies seldom are ranked accurately. In natural resources and banking there is no discernible pattern of rankings for the balance sheet. Across the five industries, of the

35 cells beyond the second actual ranking error, the majority of the auditors correctly rank error frequency plus or minus one rank for only 15 cells.

Panel B: Income Statement Accounts. To understand the confusion matrices for the income statement accounts, consider the industries for which General and Administrative Expenses are included in the income statement (i.e., manufacturing, merchandising, and natural resources): Auditors consistently underestimate the frequency of errors in General and Administrative Expenses. The matrices for merchandising and natural resources industries indicate that there is far less consensus among auditors as to the relative frequencies of the accounts (excluding General and Administrative Expenses) than is indicated for the manufacturing industry. However, about one half of the participants rank the remaining errors correctly. For insurance, most auditors correctly distinguish the two more frequently occurring from the two less frequently occurring errors. In banking, there are only two income statement accounts, and 107 (162) auditors rank them correctly (incorrectly).

Summary of the First Study

The results are as follows: (1) Regardless of their experience levels, auditors do not have very accurate knowledge of the relative frequencies with which financial statement accounts are affected by errors. (2) The accuracy of experienced auditors' knowledge about error frequencies is, on average, substantially higher for clients in the manufacturing and merchandising industries than for those in natural resources, banking, or insurance. However, auditors' accuracy in all five industries is low compared with the accuracy of frequency knowledge of a variety of events studied in the memory literature. (3) A strong positive relationship between accuracy of error-frequency knowledge and the number of months of audit experience, number of audit clients, or auditor's position in the firm does not exist for any of the five industries. (4) Auditors tend to know the one or two most frequently affected balance sheet accounts for the manufacturing, merchandising, and insurance industries, but not for the banking and natural resource industries. Also, they tend to know which income statement accounts are more frequently affected, except for General and Administrative Expenses.

The empirical results on the number of clients audited (table 3) provide a likely explanation for auditors' lack of frequency knowledge about error effects: Direct experience with a financial statement error effect is a rare event. A "frequently" occurring error in the Wright and Ashton (1989) sample affects the financial statements of, at most, 15 percent of the clients in the sample. This situation is exacerbated by the fact that, for a particular audit client, errors tend to recur in the same accounts year after year (Kinney 1979; Hylas and Ashton 1982; Wright and Ashton 1989). Auditors who have experienced only a few audits in an industry are not likely to have sufficient experience with even the most frequently occurring errors to have learned the accurate base rate frequency information through direct experience. Also, the client mix experienced by an individual auditor is unlikely to be representative of the population of audits, creating some question about the appropriate base from which error-frequency knowledge should be accumulated.

C. The Second Study

The second study examines error causes as in Libby (1985) and error effects (the task used in the first study) on a within-auditor basis. The study is conducted within auditors to compare absolute knowledge levels of error causes and effects. Participants were supervising seniors, managers, and senior managers, and most of them have manufacturing industry experience. There is some variance across auditors on the three experience variables—number of months of firm experience, number of months of manufacturing experience, and number of manufacturing clients—so the relative importance of experience on error-effect knowledge and error-cause knowledge can also be assessed within the limits of the three experience levels represented.

The comparison of error-cause and error-effect knowledge is particularly important because of the apparent discrepancy between the results of the first study, reported previously and those in Libby (1985) and Libby and Frederick (1990). The latter studies concluded that auditors know error-cause frequencies in manufacturing, and that such knowledge is positively related to experience. In contrast, the first study suggests that experience is not related to auditors' knowledge of error *effects*. According to the psychological research on which the present study is based, if auditors experience both of these aspects of errors, they should automatically acquire frequency knowledge about both.

Tasks. The elicitation technique used in Libby (1985) was used for both the error cause and error effect tasks in the second study, and the error cause task is identical to that used in Libby (1985) and Libby and Frederick (1990). The error cause task has auditors rank 12 error causes by assigning the number 100 to the most frequently occurring cause, and then assigning numbers between one and 99 to the remaining causes depending on how frequently they are thought to occur in relation to the most frequent cause. The error effect task is the same as in the first study with several modifications. (1) Since accuracy may have been adversely affected in the first study by having the auditors consider both balance sheet and income statement accounts, the balance sheet accounts are assessed separately from the income statement accounts. (2) For the balance sheet task, accounts in the "Other" category are also assessed (in addition to the 11 most frequently affected manufacturing accounts). Thus, 18 balance sheet accounts are considered separately from the four income statement accounts. (This change was made to see whether the most infrequent error events could be accurately distinguished.) (3) The elicitation technique is revised to be consistent with the error cause task, the instructions are simplified, and the format is adjusted to correspond to the error cause task as it appears in Libby and Frederick (1990).

The Accuracy Criteria. The accuracy criterion for the error cause task is errors greater than ten percent of materiality in a sample of actual audits described in Willingham and Wright (1985).[11] The criterion for the error effect task is the same as in the first study.

The Sample of Auditors. The auditors were 19 firm supervising seniors, managers, and senior managers on rotational assignment in the firm's Executive Office. They had worked for the firm between four and nine years. Sixteen auditors had experience in manufacturing, having participated in audits of up to 15 clients and having spent up to 55 months on manufacturing audits. Table 5 provides data on the individual auditors' experience.

---

[11] One error has been excluded from the data analyses of the error causes task in the prior and present studies because it was excluded from the working papers from which the criterion was taken (Libby 1985, 658). It was included in the present study only for consistency with the prior studies.

Table 5
Second Study: Accuracy and Experience in the Error Causes and Error Effects Tasks[*]

Auditor Number	Position in Firm	Mo. in PMM Practice	# Mfg. Clients	Mo. Mfg. Exper.	Error Causes Task	Error Effects Task	
						B/S Accts.	*I/S Accts.*
					Acc.	*Acc.*	*Acc.*
1	Mgr.	62	10	20	0.90	0.68	0.28
2	Mgr.	62	3	9	0.91	0.80	0.54
3	Sr. Mgr.	106	10	55	0.58	0.29	−0.09
4	Mgr.	60	3	10	0.30	0.69	0.45
5	Sr. Mgr.	77	3	15	0.70	0.53	−0.12
6	Sup. Sr.	57	7	23	0.60	0.51	0.49
7	Sup. Sr.	48	7	28	0.82	0.69	0.37
8	Mgr.	79	0	0	0.38	0.72	0.30
9	Sup. Sr.	45	6	28	0.73	0.57	0.57
10	Mgr.	64	0	0	0.68	0.14	0.29
11	Sr. Mgr.	91	15	20	0.53	0.81	0.06
12	Mgr.	67	1	3	0.45	0.70	−0.45
13	Sup. Sr.	60	6	20	0.47	0.81	0.37
14	Sr. Mgr.	54	8	24	0.84	0.43	0.23
15	Mgr.	67	0	0	0.35	0.17	−0.03
16	Mgr.	60	10	30	0.91	0.77	0.47
17	Mgr.	66	2	18	0.04	0.53	−0.27
18	Mgr.	78	10	12	0.89	0.62	0.31
19	Sr. Mgr.	96	6	32	0.45	0.58	0.36
Mean		68	6	18	0.61	0.58	0.22
Std. Dev.		16	4	14	0.25	0.20	0.29

[*] Accuracy (Acc.) is measured as the Pearson correlation between an individual's responses and the actual frequencies.

Data Collection Instrument. The seven-page collection instrument includes instructions and response forms for both tasks and a page requesting information on experience and education. One half of the auditors responded to the error-effects task first, and the other half responded to the error-causes task first.

Analysis. Three accuracy measures are calculated for each auditor—one for the error causes task, one for the error effects task for balance sheet accounts, and one for the error effects task for income statement accounts. In the error causes task, accuracy is measured as the correlation between each auditor's estimates of the relative frequencies of the 11 error types and the actual frequencies from the Willingham and Wright (1985) criterion as reported in Libby (1985, 658). Similarly, in the error effects tasks, accuracy is measured as the correlation between each auditor's estimated frequencies and the actual frequencies from the Wright and Ashton (1989) criterion. Each accuracy measure is correlated with each of three experience measures—number of months of overall audit experience, number of months of manufacturing audit experience, and number of manufacturing clients audited. In addition to these group-level analyses, individual analyses in the form of confusion matrices are presented.

Table 6
Second Study: Relation Between Manufacturing Industry Accuracy[a] and Experience in Manufacturing[b]

	Correlation Between Accuracy and Experience		
		Error Effects Task	
Experience Measure	*Error Causes Task*	*B/S Accounts*	*I/S Accounts*
Months of PMM Experience	−0.25	−0.13	−0.40
Months of Manufacturing Experience	0.21	0.01	0.11
Number of Manufacturing Clients	0.46	0.32	0.21

[a] Accuracy is measured as the Pearson correlation between an individual's responses and the actual frequencies.

[b] Recall that there is less variance in auditors' experience levels in the second study than in the first study or in Libby and Frederick (1990).

Group Results

Average individual accuracy in the error causes task is 0.61, with a range of 0.04 to 0.91, and a standard deviation of 0.25 (see table 5). Thus, the mean accuracy result for managers from Libby and Frederick (1990) ($r=0.58$) is consistent with the present results. However, experience does not explain much of the variation in individual accuracy among these experienced auditors in either task. Table 6 shows that individual accuracy correlates negatively with months of overall audit experience (−0.25), and positively correlates with months of manufacturing audit experience (0.21) and with the number of manufacturing clients audited (0.46). (Only the latter correlation is significantly different from zero at $\alpha=0.05$, df$=17$.)

The results are reported separately for balance sheet and income statement accounts in the error effects task. Table 5 shows that mean accuracy for balance sheet accounts is 0.58, with a range of 0.14 to 0.81, and a standard deviation of 0.20 for the auditors in the second study. Thus, mean accuracy is less than that found for balance sheet accounts in the first study for auditors with comparable experience (0.73 for supervising seniors and 0.74 for managers and senior managers—see table 1). For income statement accounts, table 5 shows that mean accuracy is 0.22, with a range of −0.45 to 0.57, and a standard deviation of 0.29, approximately the same accuracy found in the first study for auditors at these levels. Table 6 shows, for balance sheet (income statement) accounts, that accuracy correlates −0.13 (−0.40) with months of experience, 0.01 (0.11) with months of manufacturing experience, and 0.32 (0.21) with number of manufacturing clients. (None of these correlations is significant at $\alpha=0.05$, df$=17$.) Thus, the pattern of correlations between accuracy and experience is similar for the two error effects tasks and is similar to the industry experience results for the error causes task: The correlations between accuracy and months of overall audit experience are negative, while the correlations between accuracy and months of manufacturing experience are weakly positive, and the correlations between accuracy and number of manufacturing clients are more strongly positive, but still low.

Individual Results

Confusion matrices for error effects (panels A and B) and error causes (panel C) are shown in table 7. For the balance sheet accounts in the error-effect task (panel A), the majority of auditors correctly rank (plus or minus one rank) each of the four accounts that are most often affected by errors. (In comparing the summary percentages in tables 4 and 7, recall that sample sizes differ greatly.) However, some auditors overestimate the frequency with which the "Other" accounts are affected by errors. As in the first study, the most frequent error for the income statement accounts (panel B) is that General and Administrative Expense is not ranked sufficiently high by most auditors. The results for error causes (panel C) are similar: The majority of auditors correctly identify the two most frequent error causes, but most auditors are inaccurate in their rankings of less frequently occurring errors.

Summary of the Second Study

The group- and individual-level analyses of the second study are consistent with Libby and Frederick's (1990) results on knowledge of error causes and with results of the first study on error effects for auditors at (approximately) the manager level. However, the individual analyses of error causes knowledge suggest that auditors apparently know only the most frequently occurring causes and that low accuracy (and consensus) is associated with other errors. The absence of a strong positive relation between accuracy and three experience variables found in the error effects task in the first study for all levels of experience is replicated for the supervising seniors, managers, and senior managers in the second study for knowledge of both error effects and error causes.

III. Discussion

These results suggest that auditors do not acquire base rate error frequency knowledge through experience because their experience with actual errors is limited. This section considers two implications of the results—the need to reconsider the roles of experience and frequency knowledge in audit judgment and expertise, and the extrapolation of existing models of expert judgment and decision making to auditing.

Roles for Experience and Expertise

The results suggest that audit experience must be understood as it relates to a particular audit task. They also raise questions about the usefulness of all-encompassing experience measures, such as length of time in audit practice (see also, e.g., Bedard 1989; Davis and Solomon 1989; Bonner and Lewis 1990). Specifically, it is found that industry experience and number of clients audited vary substantially within more traditional experience measures such as length of experience and position in the firm.

Following Hasher and Zacks (1979, 1984), this study also shows that auditors do not experience errors in proportion to their population frequencies. Assuming that auditors learn the relative frequencies of errors to which they are exposed, then the present results—i.e., auditors' perceptions of error frequencies—approximate error rates in the

Table 7

Second Study: Individuals' Knowledge for the Error Effects and Error Causes Tasks[a] (n=19)

Panel A. Error Effects Task: Balance Sheet Accounts:

Assigned Rank	Actual Rank[b] 1	2	3	4	5	6	7	8	9	10	Unranked Accounts[c] A	B	C	D	E	F	G	H	Total
1	9	3	2		1	1	1				1					1			
2	4	5	6	1		1				4		1				1			
3	3	2	2	7		1				2								1	
4		5	3	6		1				2				1					
5	1		5	2	1	1	3	1	1	1	1	3	1	1	1	1		1	
6		1		2	1		1				2		1			3	1		
7	1				3		1	1	1		1	3	1	1		2	1		
8		3			1		2	3	2			1	1		3	2	3	1	
9					2	1	2	2	3	1	1	1	1		2	1	2		
10					1		1			1	1	1	1	1	1				
11							2		2		1		1		1		3	2	
12			1			2	3	2	3	2	2	2	1	2	3	2	4	4	
13	1				1	4	2	6	4	3	3	2	2	5	4	3	3	3	
14										1			1	1		1	1	1	
15					5	2	1	1	1	2		4	5	5	2	1	1	4	
16					1						1	1	2	1		1		1	
17					3	2		1	2		3		1	1	2				15
18						3		2			2							1	8
%[d]	0.68	0.53	0.58	0.79	0.10	0.05	0.21	0.32	0.26	0.11									

Panel B. Error Effects Task: Income Statement Accounts:

Assigned Rank	Actual Rank 1	2	3	4	Total
1	1	10	8	0	19
2	2	8	8	2	20
3	12	0	2	9	23
4	4	1	1	8	14

Table 7—*Continued*

Panel C. Error Causes Task[e]:

Assigned Rank	Actual Rank 1	2	3	4	5	6	7	8	9	10 =	11	12	Total
1	9	4	2	2	2	0	0	0	0	0	0	0	19
2	4	6	4	3	0	2	0	1	1	0	0	1	22
3	2	2	3	0	2	0	2	1	2	0	1	3	18
4	2	5	2	2	3	2	1	1	0	2	1	0	21
5	1	0	1	3	2	4	1	2	0	2	0	2	18
6	1	1	3	2	1	1	2	3	1	1	0	1	17
7	0	1	2	0	2	4	3	0	4	2	3	2	23
8	0	0	1	0	3	2	2	1	2	3	1	4	19
9	0	0	1	5	2	1	3	3	1	2	3	1	22
10	0	0	0	1	0	2	2	1	4	4	2	1	17
11	0	0	0	0	2	1	3	5	4	2	3	1	21
12	0	0	0	1	0	0	0	1	0	1	5	3	11
%[d]	0.68	0.63	0.47	0.26	0.32	0.47	0.37	0.21	0.37	0.42	0.42	NA[f]	228

[a] Indicates the number of auditors who assigned each rank to the account represented by the actual rank. If all auditors had ranked the accounts perfectly accurately, the diagonal cells would contain 19, and all other cells would contain zero.

[b] Actual accounts for the error effects ranks are given in table 4, footnote b.

[c] Unranked accounts are A. Cash, B. Marketable securities and other investments, C. Notes receivable, D. Long-term investments, E. Intangible assets, F. Income taxes payable, G. Notes payable and long-term debt, and H. Other long-term liabilities. These accounts are in their order of appearance in the Balance Sheet, since their actual rankings are so similar and their actual frequencies so low.

[d] Indicates the percent of total (n = 19) number of auditors who ranked the error in the correct category or in the categories one above or one below the correct category.

[e] Actual ranks for the error causes task correspond to the following: 1. Accrued expenses and payables not recorded or underrecorded; 2. Bad debt expense and allowance not recorded or underrecorded; 3. Current period purchases on account not recorded or recorded next period; 4. Next period's credit sales recorded in the current period; 5. Goods returned by customer in current period not recorded or recorded next period; 6. Expense items improperly classified as inventory; 7. Marketable securities not written down to lower of cost or market; 8. Payments on account recorded but not made or overrecording of such payments; 9. Expense payments made but not recorded or underrecording of such payments; 10. and 11. Purchase returns recorded but goods not returned or returned next period and Sales to valid customers are recorded more than once have identical frequencies; and 12. Current portion of long-term debt improperly classified as noncurrent.

[f] This error was not analyzed in the prior or present studies of the error causes task, because it was not available in the archival data. It is included here for completeness.

subsamples of audits in each auditor's experience. These results do *not* indicate whether, and to what extent, it is desirable that auditors know population archival frequencies. If such knowledge is required, then it may be learned through indirect experience or provided as needed through decision aids or external memory. Otherwise, the firm must identify auditors who possess the requisite knowledge and assign them to particular audit tasks. Determining the costs and benefits of these alternative strategies will require a rich understanding of the nature of the experience that provides error-frequency knowledge. Alternatively, auditors may not need to know error frequencies because client-specific information swamps this base rate, so that the benefit of acquiring base rate knowledge is less than the cost of doing so. Empirically, we observe that at least three accounting firms have sponsored error frequency studies, and at least one firm—KPMG Peat Marwick—has developed a data base of error frequencies for a population of international audits.

Modeling Audit Judgment/Expertise

These results—and the fact that financial statement errors apparently are rare events (see Wright and Ashton 1989)—suggest that a diagnostic model of audit judgment that emphasizes base-rate knowledge about error events (see Libby 1985) may not be adequate or appropriate to describe audit expertise. The results imply that auditors rely on knowledge about the most likely errors, rather than *overall* base rates. With hindsight, this is not surprising in a domain in which the events of interest occur so infrequently. Further, existing models of expertise rely on knowledge acquisition and retrieval from memory, based on years of frequent, direct experience with events for which a criterion that provides learning opportunities is available. These results raise questions about how, and how well, these existing models extrapolate to *audit* judgment and expertise. Future research could benefit from considering new models of audit judgment and from a deeper understanding of the institutional setting in which auditors work. The details of audit tasks and experiences should be articulated and examined vis-à-vis the incentives and constraints inherent in audit settings.

Toward this end, it is essential to consider how auditors *differ* from other experts in ways that increase the difficulty and complexity of audit research. An analogy with expert chess players may be useful for this purpose. First, unlike chess masters, expert auditors are not readily identifiable. Moreover, auditors perform a great variety of judgment (and nonjudgment) tasks, whereas the range of tasks for chess masters is limited to the game of chess. Finally, while success in chess is based on performance that can be characterized unambiguously as winning and losing, an auditor's survival and success in a professional organization depends only in part on the judgment and ability required to produce an audit opinion. In addition, the auditor must exercise interpersonal and political skills that provide client services and contribute to the maintenance and vitality of the practice. Even if one considers only *technical* expertise, it is clear that auditors as a group perform many tasks, of which any individual auditor may perform only a few. For example, the growing trend toward industry specialization and the use of statistical, computer, and other specialists suggests that all auditors are not required to exercise unaided judgment in all audit tasks.

An analogy to medicine may also be useful. The expertise of a radiologist who frequently and repetitively performs one particular type of task requires a different characterization than that of an internist or general practitioner who applies a variety of diagnostic techniques *and* provides treatment for many illnesses. The internist, in turn, may possess expertise that should be characterized differently from that of a cardiac specialist who uses many diagnostic techniques but focuses on one particular system. To the extent that the auditor whom a researcher has in mind—or uses as a subject—is a general practitioner, it should not be assumed that the auditor's expertise extends to all auditing tasks performed in all industries, in companies of varying sizes, operating in various environments, and so on.

The perspective of an auditor as a general practitioner—who is successful without being an "expert" in the specialist sense—may suggest important research avenues. For example, perhaps the less practiced an auditor is in a particular task, the more likely he or she is to use external memory or decision aids that organize information search and/or processing. The extensive use of physicians' desk references and audit manuals and checklists suggests that the essence of expertise may not be factual knowledge like error frequencies that are learned through experience and stored in the practitioner's memory. This perspective of the general practitioner versus the specialist might improve our understanding of expertise by reminding us that expertise in auditing, unlike that in many other areas, is based on phenomena that entail rare events, poorly defined criteria, and ambiguous feedback.

References

Abdolmohammadi, M. J. 1987. Decision support and expert systems in auditing: A review and research directions. *Accounting and Business Research* 17 (Spring): 173–85.

Attneave, F. 1953. Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology* 46: 81–6.

——. 1959. *Applications of Information Theory to Psychology*. New York: Holt, Rinehart, and Winston, Inc.

Bedard, J. 1989. Expertise in auditing: Myth or reality? *Accounting, Organizations and Society* 14: 113–31.

Bonner, S., and B. Lewis. 1990. Dimensions of auditor expertise. Working paper, University of Colorado, Boulder, forthcoming in *Journal of Accounting Research*.

Butt, J. 1988. Frequency judgment in an auditing-related task. *Journal of Accounting Research* 26 (Autumn): 315–30.

Chi, M. T. H., R. Glaser, and E. Rees. 1983. Expertise in problem solving. In *Advances in the Psychology of Human Intelligence*, edited by R. Sternberg. Hillsdale, NJ: Erlbaum.

Davis, J. S., and I. Solomon. 1989. Experience, expertise, and expert-performance research in public accounting. *Journal of Accounting Literature* 8: 150–64.

Einhorn, H. J. 1972. Expert measurement and mechanical combination. *Organizational Behavior and Human Performance* 7 (February): 86–106.

Elliott, R. K. 1983. Unique auditing methods: Peat Marwick International. *Auditing: A Journal of Practice & Theory* 2 (Spring): 1–12.

Hasher, L., and R. T. Zacks. 1979. Automatic and effortful processes in memory. *Journal of Experimental Psychology: General* 108 (September): 356–88.

——. 1984. Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist* 39 (December): 1372–88.

Hock, H. S., and L. Hasher. 1990. Judgments of frequency: A tool for the analysis of memory. In *Event-related Potentials of the Brain*, edited by J. W. Rohrbaugh, R. Parasuramen, and R. Johnson. New York: Oxford University Press.

Hylas, R. E., and R. H. Ashton. 1982. Audit detection of financial statement errors. *The Accounting Review* 57 (October): 751–65.

Jacoby, J., T. Troutman, A. Kuss, and D. Mazursky. 1986. Experience and expertise in complex decision making. In *Advances in Consumer Research*, edited by R. J. Lutz, 13: 469–72.

Kinney, W. R. 1975. A decision-theory approach to the sampling problem in auditing. *Journal of Accounting Research* 13 (Spring): 117–32.

———. 1979. The predictive power of limited information in preliminary analytical review: An empirical study. *Journal of Accounting Research* 17 (Supplement): 148–65.

Lesgold, A., H. Robinson, P. Feltovich, R. Glaser, D. Klopfer, and Y. Wang. 1988. Expertise in a complex skill: Diagnosing x-ray pictures. In *The Nature of Expertise*, edited by M. T. H. Chi, R. Glaser, and M. J. Farr. Hillsdale, NJ: Erlbaum.

Libby, R. 1985. Availability and the generation of hypotheses in analytical review. *Journal of Accounting Research* 23 (Autumn): 648–67.

———. 1983. Determinants of performance in accounting decisions. In *Accounting Research Convocation*, edited by K. R. Bindon. Tuscaloosa: University of Alabama.

———, and R. K. Blashfield. 1978. Performance of a composite as a function of the number of judges. *Organizational Behavior and Human Performance* 21 (April): 121–29.

———, and D. M. Frederick. 1990. Experience and the ability to explain audit findings. *Journal of Accounting Research* 22 (Autumn): 348–67.

Owens, A. 1983. Where the primary-care squeeze is hurting GPs and FPs. *Medical Economics* 60 (13 June): 104–35.

Pitz, G. 1979. Critique of procedures for the communication of uncertainty in auditor's working papers. In *Behavioral Experiments in Accounting II*, edited by T. J. Burns. Columbus: The Ohio State University.

Simon, H. A., and W. G. Chase. 1973. Skill in chess. *American Scientist* 61 (July–August): 394–403.

Willingham, J. J., and W. F. Wright. 1985. Financial statement errors and internal control judgments. *Auditing: A Journal of Practice & Theory* 5 (Fall): 57–70.

Wright, A., and R. H. Ashton. 1989. Identifying audit adjustments with attention-directing procedures. *The Accounting Review* 64 (October): 710–28.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 240–258

Auditors' Representation and Retrieval of Internal Control Knowledge

David M. Frederick
University of Colorado–Boulder

SYNOPSIS: This article examines the characteristics of experienced and inexperienced auditors' retrieval of internal controls from memory. Because it is possible that knowledge retrieval is a function of the manner in which information is stored in memory, this study also investigates two kinds of knowledge organization: (1) a taxonomic representation, as typified by categoric checklists; and (2) a schematic representation, as typified by information flowcharts. Although the relationship between knowledge and expertise is well documented in the literature, few accounting researchers have attempted to study the memory structures of skilled individuals (Libby and Frederick 1990; Weber 1980). Thus, this work is motivated by two broad objectives: (1) to understand the nature of expertise, which will help in assessing the boundaries of auditors' abilities and performance; and, (2) to comprehend the cognitive skills of both expert and novice auditors, which will help in facilitating skill acquisition by unseasoned individuals.

One hundred thirteen auditors and 97 auditing students were used to study the retrieval characteristics of the taxonomic and schematic knowledge structures and memory differences between experts and novices. Auditors freely recalled more internal controls when such controls were organized by transaction flow. Auditing students did not possess complete knowledge organizations, and the type of representation did not affect their free recall of the controls. Across a given knowledge representation, providing subjects with complete subsets of internal controls as a memory aid did adversely affect the retrieval of additional controls in the taxonomic but not the schematic organization. Within a specific class of internal controls,

I thank Sarah Bonner, Michael Johnson, William Kinney, Barry Lewis, Robert Libby, and David Meyer, as well as the seminar participants at California–Irvine, Michigan, and Virginia Tech for their helpful comments. I also thank two anonymous referees for the careful reading and detailed comments they provided. Support from the Peat Marwick Foundation Research Opportunities in Auditing program is likewise gratefully acknowledged.

Submitted June 1988.
Revisions received June 1989, June and October 1990.
Accepted October 1990.

however, providing subjects with a portion of the controls was detrimental to the recall of the remaining controls in both types of representation. These results suggest that an auditor's retrieval of internal controls from memory depends not only on the auditor's level of experience but also on the way in which the auditor's knowledge of controls is organized.

Key Words: *Long-term memory, Taxonomic and schematic representation, Cued recall, Expertise.*

Data Availability: *The data upon which this paper is based may be obtained from the author on request.*

AUDITORS work in a domain where having specific knowledge of subject-matter is crucial to successful performance. Evaluations of internal control, for example, require auditors to exercise judgment and rely on prior knowledge of errors, irregularities, and controls. This information is stored more or less permanently in long-term memory. Thus, long-term memory must be understood if the role of expertise in auditors' judgments is to be understood. Research must discover, for instance: (1) what knowledge an auditor has stored in memory, (2) how the knowledge is represented in memory, and (3) how it is accessed when needed. This article addresses these three issues in the context of a cued-recall memory task investigating auditors' internal control knowledge.

The remainder of the article is organized as follows. Section I reviews previous research on expertise and organization of knowledge. Section II discusses research relating to some known impediments to knowledge retrieval. This research is used to develop theoretical arguments on which the current study's hypotheses are based. Subsequent sections detail the experiment, results, and implications from the tests of the hypotheses. A summary and general discussion of the research conclude the article.

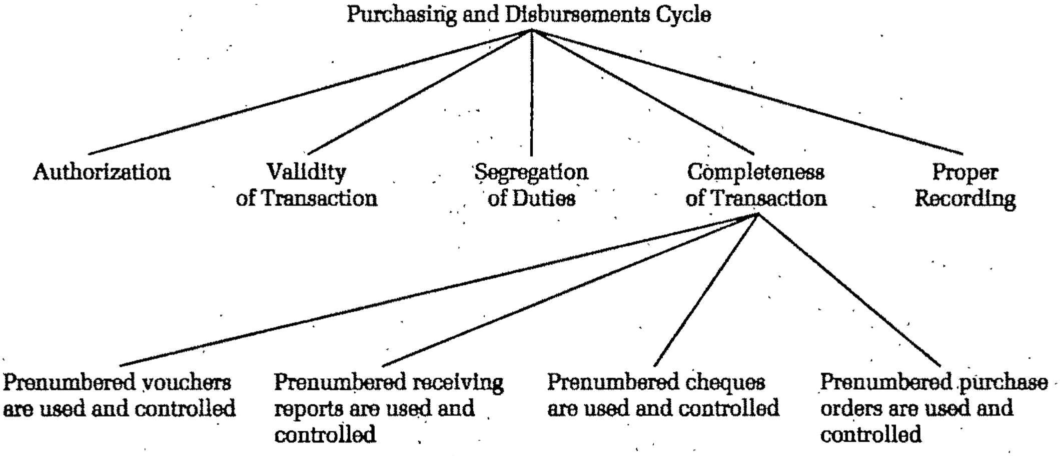
I. Organization of Knowledge in Memory

Taxonomic Organization

In general, a taxonomic organization of knowledge is a hierarchical structure in which the locations of categories are interconnected based on class membership and similarity relationships among class members (Smith 1978). The concepts forming a category tend to have attributes that co-occur only with a limited set of other attributes. For example, creatures with feathers are likely to have wings, beaks, and be able to fly, whereas creatures that have fur are not likely to have feathers, wings, or beaks, and are not likely to be able to fly (Rosch 1978; Rosch et al. 1976).

Although much of the research with taxonomic knowledge structures has focused on categorization of natural objects, reviews of auditing textbooks (e.g., Carmichael and Willingham 1989; Arens and Loebbecke 1988) and CPA firm publications reveal that this type of representation is also appropriate for internal control procedures. A business firm adopts an internal control structure to safeguard its assets and to increase the reliability of its accounting data. Thus, a firm must establish control procedures to provide assurance that the firm's detailed control objectives will be achieved. For example, control procedures must ensure *Authorization*, *Validity of Transaction*, *Segrega-*

Figure 1
Portion of Taxonomic Organization of Internal Controls


Purchasing and Disbursements Cycle
Authorization
Validity of Transaction
Segregation of Duties
Completeness of Transaction
Proper Recording
Prenumbered vouchers are used and controlled
Prenumbered receiving reports are used and controlled
Prenumbered cheques are used and controlled
Prenumbered purchase orders are used and controlled

tion of Duties, Completeness of Transaction, and *Proper Recording.* Consequently, it is argued here that internal control procedures are mentally organized around these detailed control objectives in a taxonomic representation.

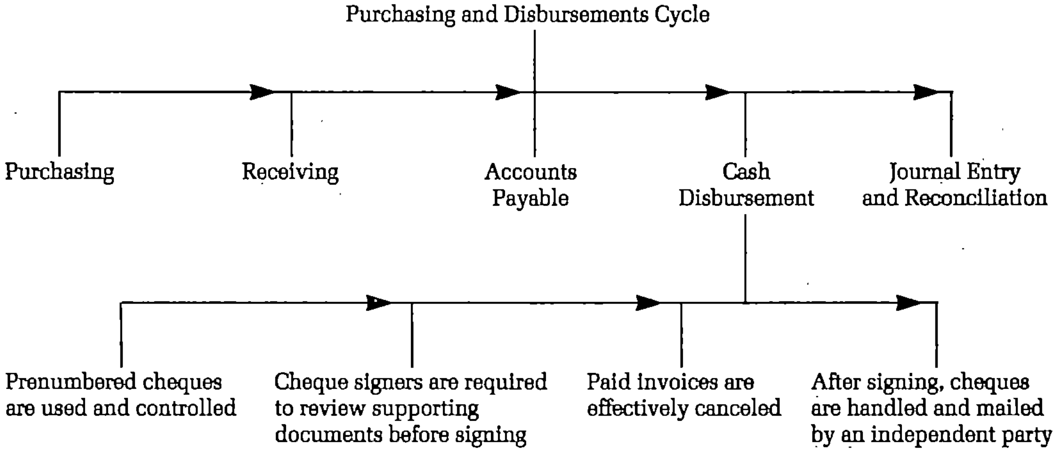
More precisely, within a taxonomic representation of internal controls, these control objectives form an intermediate level of categorization. The specific control procedures that accomplish these objectives are grouped, by objective, at a subordinate level of categorization. The accounting cycles form the superordinate level. Using the purchasing and disbursements cycle, figure 1 illustrates a portion of this taxonomic representation of internal controls.

Schematic Organization

To characterize the recall of elaborate stories and the organization of event sequences and scenes, notions like schemata (Bartlett 1932; Rumelhart and Ortony 1977; Rumelhart 1980), frames (Minsky 1975), and scripts (Abelson 1981) have been developed. While these terms are not synonymous, this paper will refer to all of the above concepts as schemata.

Similar to a taxonomic structure, a schematic representation of knowledge can be depicted as a hierarchy of components, with each component being connected to descriptions (subschemata) at a greater level of detail (Rumelhart and Ortony 1977). Unlike the taxonomic organization, however, the structure of a schema is not based on class membership and similarity relationships among class members. Rather, "it is a spatially and/or temporally organized structure in which the parts are connected on the basis of contiguities that have been experienced in space or time" (Mandler 1979, 263). Furthermore, schemata provide stereotyped descriptions about what things look like and/or the order in which they occur. As examples, activities like going to a restaurant, riding a bus, and attending a lecture are organized around schemata.

Figure 2
Portion of Schematic Organization of Internal Controls


Purchasing and Disbursements Cycle
Purchasing
Receiving
Accounts Payable
Cash Disbursement
Journal Entry and Reconciliation
Prenumbered cheques are used and controlled
Cheque signers are required to review supporting documents before signing
Paid invoices are effectively canceled
After signing, cheques are handled and mailed by an independent party

Internal control procedures can also be organized schematically. Auditors must not only be concerned with the detailed control objectives of control procedures, they must also obtain an understanding of the system for processing each significant class of transactions to evaluate internal controls. This understanding is accomplished by identifying and tracing the temporal flow of transactions within each significant transaction class. The auditing literature provides numerous examples of these sequentially ordered processes. The temporal flow of transactions within the purchasing and disbursements transaction cycle may be organized, for instance, by *Purchasing, Receiving, Accounts Payable, Cash Disbursement,* and *Journal Entry and Reconciliation.*

These sequentially aligned events form an intermediate level of activity in a schematic representation of internal controls. The specific control procedures are associated, at a subordinate level, with these events and with each other temporally as a result of the flow of transactions within a processing system. The accounting cycles continue to form the superordinate level of organization. Using the purchasing and disbursements cycle, figure 2 illustrates a portion of this schematic representation of internal controls.

Comparing Taxonomic and Schematic Organizations

Taxonomic and schematic representations of internal controls are both hierarchically arranged cognitive structures. Specific controls are grouped by, and vertically related to, control objectives in the taxonomy and activities in the schema. As a result, the two representations could be expected to influence memory in similar ways, but they differ in ways that may produce dissimilar recall patterns.

The primary difference of concern to this article arises from the nature of the organization in the schematic structure. Besides being hierarchically linked, each internal control procedure in the schema is also *temporally* connected to its neighbors (Mandler 1979). These temporal affiliations among controls may provide *more* retrieval cues or *better ordered* retrieval cues than those present in a taxonomic representation. Thus, the schematic structure may act as a more effective retrieval guide (Rabinowitz and Mandler 1983). If so, auditors' recall of internal controls should be greater when the controls are organized schematically than when they are organized taxonomically.

Using generic memory tasks, Rabinowitz and Mandler (1983) found that recall of nouns was greater when they were organized schematically than when they were organized taxonomically. However, upon reexamination of their data and design, Khan and Paivio (1988) found no difference in recall from the two organizations, suggesting that schematic and taxonomic structures provide equivalent retrieval guides. These results are inconclusive because neither study has used: (1) a *complete single set* of materials that can be structured both schematically and taxonomically; (2) a *temporally* organized schematic knowledge representation in which the stimuli are explicitly connected; or (3) a group of *participants* who are charged in their day-to-day activities with retrieving that single set of materials both by schematic cues and by taxonomic cues. The current study incorporates these features.

Expertise and Representation of Auditing Knowledge

In domains such as auditing where specific knowledge of subject-matter is necessary, experts are presumed to outperform novices. Research has revealed that experts possess a larger quantity of information and have it better organized than novices. Skilled persons are able to structure their knowledge more hierarchically, with: (1) more highly refined categories and greater differentiation at subordinate category levels, and (2) more accurate generalization at superordinate category levels (for reviews, see Chi et al. 1988; Alba and Hutchinson 1987).

Thus, it is proposed that auditing experts mentally organize internal control procedures in highly structured representations both schematically (by transaction flow) and taxonomically (by control objectives). An examination of CPA-firms' manuals reinforces this conjecture by revealing that auditors prepare workpapers using the two types of organization. Whenever auditors rely on internal controls, the controls are documented on control flowcharts *and* control objective worksheets or questionnaires. Consequently, auditors have to associate an internal control *concurrently* with a schematic and a taxonomic knowledge structure. The organizational principles of these representations suggest that auditing experts have more cues, or better ordered cues, available when retrieving internal control knowledge from a schematic structure rather than from a taxonomic structure because of the additional temporal associations among the controls.[1] Thus, the first hypothesis to be tested is:

> H1: Experts retrieve internal controls in greater quantity from a schematic organization than a taxonomic organization.

Since an auditor must know both control objectives and transaction flows, auditing novices also are exposed to both. Training materials emphasize the importance of understanding the flow of transactions within a transaction cycle and the types of internal controls that should be present. One knowledge representation is not repeatedly detailed to the exclusion of the other. Thus, novices learn to concurrently associate an internal control with a control type and a transaction cycle.

---

[1] While hypotheses described in this section are concerned with representations of auditing knowledge, the "retrieved" information is used as the dependent variable. This strategy is adopted because the amount of information recalled and the manner in which it is recalled have been demonstrated to be appropriate for assessing knowledge organization (Murphy and Puff 1982).

Research has found that, relative to experts, novices possess a smaller quantity of knowledge and that this knowledge is not highly differentiated (see Chi et al. 1988; Alba and Hutchinson 1987). Because these findings are not domain dependent, auditing novices are expected to exhibit similar characteristics. As a result of incomplete knowledge representations, auditing novices will not have available the rich temporal associations found in complete schematic knowledge representations. Thus novices will not have additional retrieval guides when retrieving information from a schematic structure. This leads to the second hypothesis:

H2: Novices retrieve internal controls equally well from schematic and taxonomic organizations.

II. Impediments to Retrieval of Knowledge

To evaluate internal controls, auditors must have both knowledge of the controls that should be present and the ability to access this information when it is needed. Instances may occur, however, when auditors are not quite as effective as they should be at retrieving available internal control knowledge. In many of these cases, the phenomenon of "output interference" in memory retrieval may be responsible.

Output Interference

Using taxonomic representations of knowledge, researchers have investigated the effects on memory recall of using category items as cues (e.g., Alba and Chattopadhyay 1985; Blaxton and Neely 1983; Roediger 1974; Rundus 1973). The typical study employs the part-list cuing paradigm in which subjects first study a list composed of multiple-item categories. After presentation of the list, participants are asked to recall all the list words they can. In the experimental conditions, subjects might be given cues of any number of the previously learned items and have to recall only the uncued items. A prevalent finding is that participants who receive multiple-item cues from each category recall a smaller proportion of the uncued items than do subjects who receive no cues. This result is termed "output interference"—the act of recalling some items inhibits recall of others.

Rundus (1973) argues that retrieving an item from memory is analogous to sampling with replacement so that previously-retrieved items may be retrieved again. Also, Rundus presumes that retrieval of an item increases the strength of that element within its category. As a result, items that have been previously retrieved from memory will tend to be retrieved again more frequently than others. As long as the retrieval and presentation of items as cues are in some sense equivalent, the part-list cuing experimental results would fit Rundus' model (for alternative models, see Nickerson 1984; Roediger and Neely 1982).

With auditors needing knowledge of internal controls to be organized both taxonomically and schematically, the output interference phenomenon allows sharpening the focus on the difference between taxonomic and schematic memory organizations. Output interference affects taxonomic structures because items within a category are not primarily connected to each other but only to the superordinate category within which they are nested. Thus, internal controls that are provided as retrieval aids do not explicitly assist recall of additional controls. In contrast, concepts within schematic structures are not only connected to their superordinate category, they are also directly connected to each other by strong temporal associations. It seems likely that providing some of these temporally related internal controls as recall cues will facilitate, not interfere with, the retrieval of others. By examining the effects of providing controls within a category as cues [part-category cuing] on both the taxonomic organization and the schematic organization of knowledge, the third and fourth hypotheses of the present research are as follows:

H3: Output interference occurs in the taxonomic representation of internal controls during part-category cuing.
H4: Output interference does not occur in the schematic representation of internal controls during part-category cuing.

The prior discussion described the effects of cuing with category instances on recall of other items from the same category. As the organizational principles of taxonomic and schematic representations apply throughout the structures, the effects of cuing with all the members of one category on the recall of instances from the remaining categories can also be investigated.

The internal controls used in this experiment belong to and are associated with the same audit area or transaction cycle because auditors usually assess the control adequacy of a cycle taken as a whole. If organized taxonomically, this is the only explicit way the controls are affiliated. They are not directly related to each other. Therefore, providing entire categories of internal controls as cues (category cuing) should produce interference in a taxonomic structure during the recall of instances from an uncued category. In a schematic or temporally organized representation, each internal control is additionally connected to its neighbors, and providing whole categories of internal controls as cues ought to facilitate retrieval of instances from an uncued category.[2] Thus, the fifth and sixth hypotheses to be tested are:

H5: Output interference occurs in the taxonomic representation of internal controls during category cuing.
H6: Output interference does not occur in the schematic representation of internal controls during category cuing.

III. Research Method

Subjects

The subjects classified as experts were 113 auditors from one of the six largest CPA firms with at least three, and an average of four, years of audit experience.[3] The novice participants were 97 undergraduate accounting majors with an average of 24 semester credit hours of accounting. The experiment was administered to the students near the end of the term during one of their normal auditing class periods. In a similar setting, the auditors completed the experiment during one of their national training seminars.

---

[2] No differences have been posited between experts and novices for hypotheses H3 through H6. However, if the free recalls provide evidence of novices being *unable* to organize knowledge schematically, then hypotheses H4 and H6 will not be applicable to them.

[3] An auditor who has gained at least three years of audit experience is one who typically has become a certified public accountant, has been in charge of an audit, and has become proficient at internal control evaluation (see Abdolmohammadi and Wright 1987).

General Case Construction

The subjects received a booklet containing an introduction, general instructions, the task to be completed, and a short debriefing questionnaire.[4] The introduction informed participants that the experiment was part of a project aimed at examining internal control structures and the judgments made about these structures by auditors. Confidentiality of responses and the importance of independent work were stressed at this point. To make the experiment less abstract, background data typical of that prepared during an audit was included as part of the general instructions. Subjects were advised that they were on the audit of a new client and that background information was being presented to familiarize them with the type of client (manufacturing) considered in the study. The debriefing questionnaire requested information about the participant's background and reaction to the experiment.

Materials

Through a review of the auditing literature, a total of 41 internal control procedures for the purchasing and disbursements cycle were identified for potential use. They were then grouped as belonging to one of five categories both schematically and taxonomically (see figs. 1 and 2). The controls were organized schematically by transaction flow as: *purchasing* [S-1]; *receiving* [S-2]; *accounts payable* [S-3]; *cash disbursement* [S-4]; and *journal entry and reconciliation* [S-5]. Taxonomically, the same controls were grouped by control objective as: *authorization* [T-1]; *validity of transaction* [T-2]; *segregation of duties* [T-3]; *completeness of transaction* [T-4]; and *proper recording* [T-5]. Once sorted by categories, the 41 potentially usable internal controls were reduced[5] to a working set of 33.

In the schematic condition, the internal controls were sequentially ordered in accordance with the typical transaction flow of the purchasing and disbursements cycle. This flow of transactions establishes the temporal ordering of the internal controls as auditors are especially concerned with the controls over each step in this flow. Flowchart exhibits that are illustrated in CPA firm publications and auditing textbooks were consulted to ensure a proper sequencing of the stimuli. These sources are in virtually unanimous agreement as to the typical transaction flow through the purchasing and disbursements cycle.

Four critical controls (those that never served as cues) were then identified by imposing the following constraints. First, to minimize the effects of primacy and recency on recall, the beginning or ending portions of the control sequence were excluded. Thus, two controls had to come from S-2 and two from S-4. Next, the four controls had to be counterbalanced across the organizations. That is, for each control in S-2 and S-4, one had to belong to T-2 and one to T-4. The controls' serial positions in the listings also had to be the same across organizations. The four controls satisfying the above constraints were: *goods received are physically verified; prenumbered receiving reports are used and controlled; prenumbered cheques are used and controlled;* and *cheque signers are required to review supporting documents before signing.*

---

[4] Pilot testing was performed to ensure that the task booklet was self-contained. It was desired to have all participants independently perform the task so that experimenter expectancies could not be conveyed to the subjects during the administration of the experiment.

[5] Of the eight controls not used, four were omitted because the professional literature did not agree as to their taxonomic grouping. The other four were randomly selected for exclusion to balance category size between the two types of representation.

As both category size and the critical controls' serial positions (positions 10, 11, 23, and 24) were the same across organizations, the presentation order of the internal controls in the taxonomic condition was constrained. However, within each taxonomic category, the presentation positions of the controls were randomized.

Information regarding internal controls is typically documented on an organizing workpaper during an audit, and the internal controls used in this experiment were worded in the same manner. The appendix shows the derived taxonomic and schematic listings of controls presented to the participants.

Recall Procedure

Subjects were randomly assigned to one of six experimental conditions formed by combining two levels of stimulus organization (schematic or taxonomic) with three levels of recall method, one of which was free recall where the participants were not provided with cues and were asked to list all 33 internal controls. The remaining two recall conditions were cued recall presentations. Under both conditions, participants were cued with all of the controls from categories S[T]-1, S[T]-3, and S[T]-5. However, the two cued recall presentations differed in which category S[T]-2 or S[T]-4 was *part-category* cued. Subjects were *either* cued with zero controls from S[T]-2 and the five noncritical controls from S[T]-4 *or* with the five noncritical controls from S[T]-2 and zero controls from S[T]-4. This manipulation resulted in two cued recall presentations of 24 internal controls each.[6] The 24 cues were presented to the subjects in the same relative order as they had been in the studied (see below) listing of controls.

The recall task proceeded as follows. All the participants first read the introduction and general instructions sections of the task booklet. They were then informed that they would be presented with a set of internal controls typical of the purchasing and disbursements cycle and would later be asked to recall these controls. All the subjects then turned the booklet to the page where the 33 internal controls were listed and studied the listing for five minutes.[7] At the end of the five-minute study period, the subjects read silently, while the experimenter read aloud, further instructions, as a distractor task in order to minimize primacy and recency effects on recall. The instructions stated that the participants were to recall the controls "in the order in which they come to mind." A lined sheet was provided to them for their responses. The cued recall conditions had the 24 cues listed at the top of this sheet (for comments regarding these procedures, see Murphy and Puff 1982). While there was no time limit set for recalling the controls, all subjects had completed the experiment within 30 minutes, taking approximately 15 minutes to recall the set of internal controls.

---

[6] The two other possible combinations of cues from the critical items' categories [zero controls from both—noncritical controls from both] were not used as they would have resulted in inconsistent numbers of cues being presented during cued recall.

[7] The subjects saw the listing of internal controls, exclusive of headings and without the critical controls highlighted, as presented in the appendix. Pilot testing determined that five minutes was an appropriate amount of time to read the 33 internal control statements. This amount of time allowed each statement to be read, and studied, for nine seconds.

Sorting Procedure

To check the validity of the experimental listings of controls, 40 auditors who did not perform the recall task were given the 33 internal controls to sort. These auditors were employed by the same firm as those who completed the recall task and had an average of five years of audit experience. One randomized order was generated with the constraint that no two controls from either a schematic or taxonomic category be adjacent. Subjects were instructed that they would be given a packet of cards on each of which was written a particular internal control; their task would be to sort the cards into piles according to one of three conditions: sorting the controls into five categories grouped by "objective function" (taxonomic-sort), by "transaction flow" (schematic-sort), or by controls that "go together in some way" (free-sort). No category names were given to any of the participants. Ten subjects were in the taxonomic-sorting group, ten in the schematic-sorting group, and 20 in the free-sorting group. Once the internal controls were sorted into groups, all participants were asked to write a short title for each pile of controls.

IV. Results

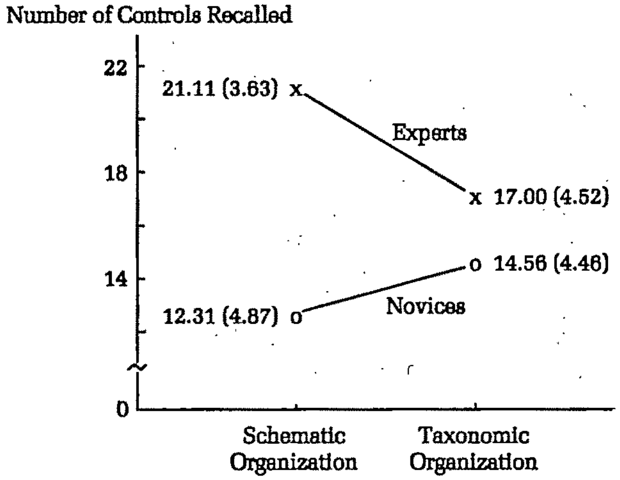
Organization of Knowledge

Internal Controls Sorting Task. The purposes of the sorting task were: (1) to provide evidence of experts' preferences for the categorization schemes; (2) to verify category titles; and (3) to verify category assignment. First, the free-sort condition of the task was used to determine the experts' preferences for the categorization schemes. This determination is important because the auditors' free-recall performances might be affected by differential familiarity (i.e., of using one knowledge organization and not the other organization during the conduct of their audit work). If so, when provided with a choice as how to organize the controls, the subjects should consistently prefer one of the categorization schemes. Analysis of the free-sort condition revealed the participants did *not* prefer one organization over the other. Within this condition, nine of the experts sorted the controls schematically, while 11 sorted the controls taxonomically.

Verification of category titles was accomplished by having the participants label each grouping of controls upon completion of the sorting task. The five groupings formed during the task were labeled with the same title by 100, 95, 95, 100, and 84 percent, respectively, of the subjects who sorted schematically, and by 95, 86, 100, 90, and 86 percent, respectively, of those who sorted taxonomically. Thus, for both knowledge organizations, experts agreed on the titles of the intermediate-level categories to which specific controls are related. These category titles were identical to the ones used when compiling the experimental materials (see figs. 1 and 2).

Finally, category assignment was verified by noting the instances when internal controls were inconsistently categorized. Interestingly, in both the free-sort and directed-sort conditions, subjects who sorted schematically misclassified, on average, fewer internal controls than those who sorted taxonomically. The mean number of schematic and taxonomic misclassifications were 5.8 versus 9.8, respectively, in the free-sort condition, and 4.3 versus 8.3, respectively, in the directed-sort conditions. A one-way analysis of variance [ANOVA] test of the combined means disclosed they were different, $F=32.8$, $p<0.001$. That fewer misclassifications occurred in the schematic

Figure 3
Number (Standard Deviation) of Controls Freely Recalled



Factor	*df*	*SS*	*F-stat.*	*Signif.*
Expertise	1, 66	548	28.71	0.0000
Organization	1, 66	15	0.78	0.3793
Expertise × Organization	1, 66	175	9.19	0.0035
Residual		1,259		

versus taxonomic condition may be suggestive of the systematic interconnections among internal controls in the schema.

Number Freely Recalled. Figure 3 shows the mean number of internal controls freely recalled by both the expert and novice subjects from the schematic and the taxonomic conditions. A 2×2 (Expertise×Organization) ANOVA was performed on the number of controls recalled. The analysis revealed significant effects for expertise, $F(1,66)=28.71$, $p<0.001$, and for the interaction between expertise and organization, $F(1,66)=9.19$, $p<0.005$. Consistent with the notion that experts possess a great quantity of knowledge, auditors recalled more internal controls than did novices. Figure 3 also indicates that hypotheses H1 and H2 are not contradicted as the interaction between expertise and organization was primarily due to auditors' recalling more controls in the schematic than the taxonomic condition. Supplementary analysis of the simple main effects of organization strengthens this perception. The difference between organizations in amount recalled was significant for the experts, $F(1,66)=8.39$, $p<0.01$, but not for the novices, $F(1,66)=2.12$, $p>0.14$ (Winer 1971, 347).

Structure of Free Recalls. In addition to assessing the quantity of internal control knowledge retrieved (i.e., hypotheses H1 and H2), the effect of the organizational manipulation on the manner in which the participants structured their recall was examined. This ancillary analysis regarding the structure of the free recalls is important because an appropriate test of the schematic knowledge organization depends on the stimuli having strong temporal associations. These temporal affiliations are hypothesized to differentiate schematic representations with more powerful sets of retrieval cues. Adjusted ratio of clustering scores and Spearman rank correlations were used to examine the structure of the free recalls.

First, the adjusted ratio of clustering (ARC) measure was computed to determine the amount of organization present in the subjects' lists of controls.[8] Two ARC scores, one based on the schematic organization and one based on the taxonomic organization, were computed for *each* subject and were compared using a mixed ANOVA. The between-subjects factors were expertise and organization. The within-subjects factor was basis of ARC score (schematic or taxonomic).[9] Between-subjects, the mean ARC scores for the experts and novices, respectively, were 0.592 and 0.433 for the schematic treatment and 0.325 and 0.271 for the taxonomic treatment.

The analysis revealed an ARC score basis by organization interaction $F(1,66) = 24.75$, $p < 0.001$, indicating that participants clustered their recall in accordance with the assigned organization treatment condition.[10] Furthermore, subjects in the schematic organization condition clustered their output more than those in the taxonomic condition, $F(1,66) = 7.47$, $p < 0.01$. Comparable to number of controls recalled, there was an effect of expertise on structure as experts clustered their recall lists more than novices did, $F(1,66) = 9.95$, $p < 0.005$. However, no interaction between expertise and organization was found, $F(1,66) = 0.15$, $p > 0.70$, implying that the organization of the stimuli did not have a differential effect on experts versus novices.

Next, to ascertain whether the internal controls were sequentially recalled in accordance with their presentation order, Spearman rank correlations were computed and tested in a 2×2 (Expertise × Organization) ANOVA. This test was undertaken primarily to find out if auditors structure internal control knowledge with temporal associations analogous to those present in the schematic organization treatment. The mean rank correlations for the experts and novices, respectively, were 0.681 and 0.547 for the schematic condition and 0.234 and 0.212 for the taxonomic condition.[11] As these correlations denote, both experts and novices order their output to a greater extent in the schematic condition, $F(1,66) = 20.32$, $p < 0.001$. Similar to the cluster analysis, there was no significant interaction between expertise and organization, $F(1,66) = 0.42$, $p > 0.52$. In contrast to the previous analysis, there was also no effect for expertise, $F(1,66) = 0.81$, $p > 0.37$. Thus, these results suggest that participants perceived the schematic treatment as representing an actual, temporally associated flow of transactions or events. Even the novices listed the few controls that they have available in a coherent sequence.[12]

[8] The ARC score denotes the number of category item repetitions, above that expected by chance, present in a subject's recall compared to the maximum number of repetitions possible above chance. An ARC score has a fixed maximum value of plus or minus one and a fixed chance value of zero. The measure is appropriate for this experiment as it does not depend on the quantity of items recalled (Murphy 1979).

[9] Portions of the schematic [taxonomic] listing presented internal controls clustered together that belonged to the same taxonomic [schematic] grouping. Computing two ARC scores is necessary as these controls could potentially cue participants to use either organization during the recall task. Note that the sequential ordering of controls in the schematic condition is solely determined by the temporal flow of transactions and cannot be altered.

[10] Six experts (four in the schematic condition) and four novices (three in the schematic condition) did not recall according to their organization condition. The opposite treatment was the basis for the recall in all ten cases.

[11] Eliminating the four experts in the schematic condition who recalled taxonomically would increase the experts' schematic mean rank correlation from 0.681 to 0.816. Similarly, eliminating the three novices would increase their mean rank correlation from 0.547 to 0.602.

[12] In an attempt to determine if the novices' low, yet seemingly structured, outputs were due to primacy or recency effects, analysis of the percentage of controls listed per category was undertaken. Pairwise *t*-tests of the

Table 1

Percentage (Standard Deviation) of Free- and Cued-Recall of Critical Controls

		Free Recall	*Cued Recall*
Experts	Schematic Condition	0.711 (0.239)	0.568 (0.247)
	Taxonomic Condition	0.697 (0.229)	0.513 (0.228)
Novices	Schematic Condition	0.438 (0.266)	0.375 (0.238)
	Taxonomic Condition	0.547 (0.262)	0.359 (0.245)

Factor	*df*	SS	F-*stat.*	*Signif.*
E	1, 200	1.711	29.09	0.0000
O	1, 200	0.002	0.03	0.8528
R	1, 200	0.959	16.31	0.0001
E×O	1, 200	0.075	1.27	0.2611
E×R	1, 200	0.017	0.29	0.5913
O×R	1, 200	0.079	1.35	0.2468
E×O×R	1, 200	0.020	0.35	0.5566
Residual		11.762		

E = Expertise
O = Organization
R = Recall Condition

Output Interference

Hypotheses H3 through H6 propose that output interference will occur in the taxonomic organization condition but not in the schematic organization condition during both part-category and category cuing. Thus, output interference is investigated initially by comparing the internal control listings of subjects in the free-recall condition with those of cued-recall. Proportions of the four critical controls recalled from categories S[T]-2 and S[T]-4 were calculated and tested in a 2×2×2 (Expertise×Organization×Recall Condition) ANOVA.[13] Table 1 presents the mean proportions that were recalled. Experts listed a higher percentage of the critical items than did novices, $F(1{,}200)=29.09$, $p<0.001$. However, the interaction between organization and recall condition was not significant, $F(1{,}200)=1.35$, $p>0.24$, which is inconsistent with expectations. The main effect of recall condition, $F(1{,}200)=16.31$, $p<0.001$, provides further evidence that output interference occurs within both knowledge organizations.

Category Cuing. The effects of cuing with whole categories on the recall of controls from uncued categories [category cuing] were determined, within knowledge organiza-

---

mean difference in recall percentages were computed for all pairs of categories. No significant differences were found. The apparent structure of the novices' (and also the experts') output, particularly in the schematic condition, does not seem to be the result of any primacy or recency effect.

[13] Proportions of critical items recalled are used in all the analyses testing hypotheses H3 through H6. As a result, the homogeneity of variances assumption underlying ANOVA is violated. Several comments justifying the use of proportions are relevant. First, in tests of category cuing, the dependent variable needs to be normalized because free-recall subjects had an opportunity to recall a greater number of controls. Second, the F-test is relatively unaffected by heterogeneity of variances (Kirk 1982). Last, reanalysis of the data using the arcsin square root transformation (Press 1972) did not result in any changes in the substantive findings.

tion, by examining recall proportions for those instances when S[T]-2 and S[T]-4 *were not* part-category cued. As indicated earlier, cued condition subjects had to recall controls both when categories were and were not part-category cued. These participants had to recall the two critical controls within either S[T]-2 or S[T]-4, and all seven controls from the other category (S[T]-4 or S[T]-2). To test for the effects of category cuing, the percentage of the seven controls listed from the category that was not part-category cued was computed for each cued-recall condition subject. Similarly, for each participant in the free-recall condition, the percentages of these controls listed from both categories (S[T]-2 and S[T]-4) were calculated.

Separate ANOVAs, one for each knowledge organization, each with the factors of expertise and recall condition, were used to test for differences in recall proportions.[14] Within the *taxonomic* organization, the combined (i.e., T-2 analyzed jointly with T-4) category recall percentages in the free-recall condition versus the category cued condition, respectively, were 0.571 versus 0.410 for the experts and 0.491 versus 0.286 for the novices. Both groups appear affected by output interference, $F(1,101)=17.36$, $p<0.001$, while experts again recalled a higher proportion of controls, $F(1,101)=5.39$, $p<0.05$. These results also held in both categories (T-2 and T-4) when analyzed individually, although at different levels of statistical significance (category T-2: recall condition at $p<0.001$, expertise at $p<0.05$; category T-4: recall condition at $p<0.01$, expertise at $p<0.11$). Hypothesis H5 is thus supported. Output interference is brought about in the taxonomic condition for both experts and novices by category cuing.

The combined category proportions recalled within the *schematic* organization in the free-recall condition versus the category cued condition, respectively, were 0.587 versus 0.564 for the experts and 0.393 versus 0.351 for the novices. Expertise had the only reliable effect, $F(1,101)=21.73$, $p<0.001$. Output interference did not occur, $F(1,101)=0.55$, $p>0.45$, thus supporting hypothesis H6. These results also held in both categories (S-2 and S-4) when analyzed individually, although again at different levels of statistical significance (category S-2: recall condition at $p>0.17$, expertise at $p<0.001$; category S-4: recall condition at $p>0.89$, expertise at $p<0.11$). These results suggest that the output interference found initially with schematically organized knowledge is *not* brought about by *category* cuing for either experts or novices.[15]

Part-Category Cuing. Finally, the effects of part-category cuing were determined, both within and between knowledge organization, by comparing the recall proportions obtained when S[T]-2 and S[T]-4 were, and were not, part-category cued. All subjects in the two cued recall groups were part-category cued for one of two categories and category cued for the other. Thus, the type of cue is a *within-subjects* factor in this analysis.

Table 2 displays the mean percentages of the four critical internal controls recalled by the part-list cued subjects when category and part-category cued. Included in the

---

[14] Hypotheses H5 and H6 relate to the recall of internal controls from uncued categories. Separate ANOVAs for each knowledge organization must be used because the categories S-2 and S-4 contain different controls than do T-2 and T-4. Differential familiarity, if any, with the controls between the schematic and taxonomic organization would confound the results if a single ANOVA with knowledge organization as another factor was used.

[15] For each subject, the percentage (0, 50, 100) of the two critical controls recalled from the category not part-category cued was computed. The findings were similar to those for the seven total controls recalled from each category. Within the taxonomic condition, expertise ($p<0.05$) and recall condition ($p<0.005$) were both significant. Within the schematic condition, expertise ($p<0.001$) was significant while recall condition was not ($p>0.78$).

Table 2

Percentage (Standard Deviation) of Critical Controls Recalled

		Free Recall	*Category Cued**	*Part-Category Cued**
Experts	Schematic Condition	0.711 (0.239)	0.703 (0.343)	0.432 (0.357)
	Taxonomic Condition	0.697 (0.229)	0.500 (0.349)	0.527 (0.332)
Novices	Schematic Condition	0.438 (0.266)	0.422 (0.362)	0.328 (0.326)
	Taxonomic Condition	0.547 (0.262)	0.344 (0.322)	0.375 (0.311)

Factor	*df*	*SS*	*F-stat.*	*Signif.*
E	1, 134	2.063	17.97	0.0000
O	1, 134	0.083	0.73	0.3958
E×O	1, 134	0.025	0.22	0.6392
C	1, 134	0.401	3.47	0.0646
C×E	1, 134	0.140	1.21	0.2727
C×O	1, 134	0.765	6.62	0.0112
C×E×O	1, 134	0.127	1.10	0.2956
Residual		15.479		

* Note that this type of cue is a within-subjects factor.
E = Expertise
O = Organization
C = Type of Cue

table as reference points, but not in the analysis, are the free-recall subjects' recall percentages of these controls. The cued participants' critical controls recall was tested by an ANOVA in which the between-subjects factors were expertise and organization and the within-subjects factor was type of cue, category versus part-category. As before, experts outperformed novices $F(1{,}134)=17.97$, $p<0.001$. This analysis shows that output interference results from category and part-category cuing in the *taxonomic* organization, supporting both hypotheses H3 and H5. However, in the *schematic* organization only part-category cuing produced output interference. Note that the type of cue interacted with organization, $F(1{,}134)=6.62$, $p<0.05$, and recall with category cuing was marginally higher than with part-category cuing, $F(1{,}134)=3.47$, $p<0.07$. Output interference was not expected in the schematic condition during either part-category (hypothesis H4) or category (hypothesis H6) cuing. Thus, hypothesis H4 is not supported by these results.

As supplementary analysis, ANOVAs with a between-subjects factor of organization and a within-subjects factor of type of cue were performed on each expertise level. This analysis revealed that the above finding is due largely to the performance of the experts. They recalled more when category cued than when part-category cued, $F(1{,}72)=4.31$, $p<0.05$, and they had the type of cue interact with organization, $F(1{,}72)=6.44$, $p<0.05$. With regard to novices, no statistical difference existed between type of cue, $F(1{,}62)=0.31$, $p>0.58$, nor did type of cue interact with organization, $F(1{,}62)=1.22$, $p>0.27$.

V. Discussion and Conclusions

This experiment was performed to investigate how participants with different levels of expertise mentally represent internal control knowledge, and whether these representations differentially affect information retrieval. Experienced auditors freely recalled more internal control procedures from the schematic than the taxonomic organization. Apparently, the temporal affiliations among controls in a schema provide this structure with more effective retrieval guides than are present in a taxonomy.

It is possible that auditors have a differential familiarity with the representations, which may produce two effects: (1) a preference for schematically organized internal control knowledge, which is not borne out by the sorting task results; or, (2) a greater quantity of controls being accumulated (stored) in the schematic representation. If auditors use schematically organized knowledge to guide their investigative work and taxonomically organized structures to assimilate their findings, then the second effect may occur when auditors experience only investigative work. Thus, to mitigate against this bias, the current experimental task was completed by subjects who are experienced at performing both evidence collection and evaluation (Abdolmohammadi and Wright 1987). Nevertheless, this argument cannot be completely discounted as the current study does not test specifically for disproportionate amounts of knowledge being in the knowledge representations based upon level of audit experience.

A different issue is that internal control knowledge not used as stimuli could adversely affect recall in the taxonomic condition. Controls pertaining to the same objective that belong to *other transaction cycles* are stored within the taxonomic structure as category members with purchasing cycle controls (e.g., sales cycle controls). Overlap occurs because every transaction cycle must have certain detailed control objectives (e.g., segregation of duties). In contrast, within the schematic structure, transaction flows and the controls pertaining to them do not overlap. These differences could cause a potentially more coherent grouping of controls to be used in the schematic condition than in the taxonomic condition.

A count of the number of intrusions (the controls recalled that were not part of the stimuli) by the auditors provides some evidence of the coherence of the two knowledge structures. Of the controls recalled from the taxonomic and the schematic organizations, 6.4 and 2.9, respectively, were intrusions. Close to one-half of the intrusions in the taxonomic structure were between-cycle intrusions (controls belonging to cycles other than purchasing). No intrusions in the schematic structure were between-cycle intrusions. Although representing a small proportion of total responses, the occurrence of between-cycle intrusions in the taxonomic, but not schematic, condition supports the different principles of organization. This result also suggests that, in the taxonomic knowledge structure alone, controls external to the stimuli could interfere with, and thus prevent, recall of the target controls.

Besides examining how the different knowledge organizations affect free recall, this paper also investigated their effects on cued recall. It was expected that a taxonomic representation would promote output interference and that a schematic representation would not. Output interference did occur in the taxonomic structure when the auditors were both category and part-category cued. However, output interference was inhibited by the schematic representation only when the auditors were category cued. When they were part-category cued, output interference was evident.

A possible explanation for this unexpected result centers on whether it was caused by the design of the task. The cued-recall subjects were simultaneously "category" and "part-category" cued. If these participants identified the missing "category" first, they may have felt that their task was finished, as only two internal controls would have been left to be generated. To address this concern, future experiments could: (1) use part-category cuing exclusively, (2) vary the proportion of controls that must be recalled, and (3) vary the size of the presented set of controls.

Presumably, experienced auditors can perform some judgment tasks more effectively than can novices. As exhibited here by their retrieval of internal controls from memory, this superiority may be primarily a matter of the task-specific knowledge that the experts can bring to bear upon audit decisions. Thus, the continued study of skilled auditors' long-term memories is essential for both assessing their abilities and improving their performance.

Appendix

Taxonomic Listing of Internal Controls

Purchase requisitions are properly authorized.
Purchase orders are properly authorized.
There is an approved list of suppliers.
Vouchers are properly authorized.
Cheque signers are authorized by board of directors.
There are written authorization procedures and limits of buying power.
Spoiled cheques are mutilated and kept on file.
Invoices are matched to receiving reports.
Paid invoices are effectively cancelled.
GOODS RECEIVED ARE PHYSICALLY VERIFIED.
CHEQUE SIGNERS ARE REQUIRED TO REVIEW SUPPORTING DOCUMENTS BEFORE SIGNING.
Access to receiving areas is adequately controlled.
Receiving reports are sent directly to accounts payable department.
There is a separate and independent accounts payable department.
Receiving reports are checked for description and quantities by the department goods are routed to.
There is a separate and independent purchasing department.
Invoices are received by accounts payable department directly from mail opener.
There is a separate and independent cash disbursements department.
After signing, cheques are handled and mailed by an independent party.
There is a separate and independent receiving department.
There is adequate control over direct shipments from suppliers to customers.
Prenumbered vouchers are used and controlled.
PRENUMBERED RECEIVING REPORTS ARE USED AND CONTROLLED.
PRENUMBERED CHEQUES ARE USED AND CONTROLLED.
Receiving reports are matched to purchase orders.
Prenumbered purchase orders are used and controlled.
Assets are periodically inspected and reconciled to the records.
Regular summaries of journal entries are prepared and reviewed to ensure all standard entries have been made.
There are regular comparisons of expense accounts to budgets.
Batch totals of invoices are checked to totals of invoices recorded.
There are regular reconciliations of the accounts payable subsidiary ledger to the general ledger.
Invoices are checked for prices, quantities, and mathematical accuracy.
Journal entries are standardized.

Appendix—*Continued*

Schematic Listing of Internal Controls

Purchase requisitions are properly authorized.
There is a separate and independent purchasing department.
There is an approved list of suppliers.
There are written authorization procedures and limits of buying power.
Prenumbered purchase orders are used and controlled.
Purchase orders are properly authorized.
There is a separate and independent receiving department.
Access to receiving areas is adequately controlled.
There is adequate control over direct shipments from suppliers to customers.
GOODS RECEIVED ARE PHYSICALLY VERIFIED.
PRENUMBERED RECEIVING REPORTS ARE USED AND CONTROLLED.
Receiving reports are checked for description and quantities by the department goods are routed to.
Receiving reports are sent directly to accounts payable department.
There is a separate and independent accounts payable department.
Invoices are received by accounts payable department directly from mail opener.
Invoices are checked for prices, quantities, and mathematical accuracy.
Invoices are matched to receiving reports.
Receiving reports are matched to purchase orders.
Prenumbered vouchers are used and controlled.
Vouchers are properly authorized.
There is a separate and independent cash disbursements department.
Cheque signers are authorized by board of directors.
PRENUMBERED CHEQUES ARE USED AND CONTROLLED.
CHEQUE SIGNERS ARE REQUIRED TO REVIEW SUPPORTING DOCUMENTS BEFORE SIGNING.
Paid invoices are effectively cancelled.
After signing, cheques are handled and mailed by an independent party.
Spoiled cheques are mutilated and kept on file.
Journal entries are standardized.
Batch totals of invoices are checked to totals of invoices recorded.
There are regular reconciliations of the accounts payable subsidiary ledger to the general ledger.
Regular summaries of journal entries are prepared and reviewed to ensure all standard entries have been made.
There are regular comparisons of expense accounts to budgets.
Assets are periodically inspected and reconciled to the records.

References

Abdolmohammadi, M., and A. Wright. 1987. An examination of the effects of experience and task complexity on audit judgments. *The Accounting Review* 62 (January):1–13.

Abelson, R. P. 1981. Psychological status of the script concept. *American Psychologist* 36 (July): 715–29.

Alba, J. W., and A. Chattopadhyay. 1985. Effects of context and part-category cues on recall of competing brands. *Journal of Marketing Research* 22 (August): 340–49.

——, and J. W. Hutchinson. 1987. Dimensions of consumer expertise. *Journal of Consumer Research* 13 (March): 411–54.

Arens, A. A., and J. K. Loebbecke. 1988. *Auditing: An Integrated Approach*. 4th ed. Englewood Cliffs, NJ: Prentice-Hall.

Bartlett, F. C. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge, United Kingdom: Cambridge University Press.

Blaxton, T. A., and J. H. Neely. 1983. Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory and Cognition* 11 (September): 500–10.

Carmichael, D. R., and J. J. Willingham. 1989. *Auditing Concepts and Methods: A Guide to Current Auditing Theory and Practice*. 5th ed. New York: McGraw-Hill.

Chi, M. T. H., R. Glaser, and M. J. Farr. 1988. *The Nature of Expertise*. Hillsdale, NJ: Lawrence Erlbaum.

Khan, M., and A. Paivio. 1988. Memory for schematic and categorical information: A replication and extension of Rabinowitz and Mandler (1983). *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14 (July): 558–61.

Kirk, R. E. 1982. *Experimental Design: Procedures for the Behavioral Sciences.* Monterey, CA: Brooks/Cole.

Libby, R., and D. M. Frederick. 1990. Expertise and the ability to explain audit findings. *Journal of Accounting Research* 28 (Autumn): 348–67.

Mandler, J. M. 1979. Categorical and schematic organization in memory. In *Memory Organization and Structure*, edited by R. Puff, 259–99. New York: Academic Press.

Minsky, M. 1975. A framework for representing knowledge. In *The Psychology of Computer Vision*, edited by P. H. Winston, 211–77. New York: McGraw-Hill.

Murphy, M. D. 1979. Measurement of category clustering in free recall. In *Memory Organization and Structure*, edited by R. Puff, 51–83. New York: Academic Press.

——, and C. R. Puff. 1982. Free recall: Basic methodology and analyses. In *Handbook of Research Methods in Human Memory and Cognition*, edited by C. R. Puff, 99–128. New York: Academic Press.

Nickerson, R. S. 1984. Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory and Cognition* 12 (November): 531–52.

Press, S. J. 1972. *Applied Multivariate Analysis.* New York: Holt, Rinehart, and Winston.

Rabinowitz, M., and J. M. Mandler. 1983. Organization and information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9 (July): 430–39.

Roediger, H. L. 1974. Inhibiting effects of recall. *Memory and Cognition* 2 (April): 261–69.

——, and J. H. Neely. 1982. Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology* 36 (June): 213–42.

Rosch, E. 1978. Principles of categorization. In *Cognition and Categorization*, edited by E. Rosch and B. B. Loyd, 28–48. Hillsdale, NJ: Lawrence Erlbaum.

——, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8 (July): 382–439.

Rumelhart, D. E. 1980. Schemata: The building blocks of cognition. In *Theoretical Issues in Reading Comprehension*, edited by R. J. Spiro, B. C. Bruce, and W. F. Brewer, 33–58. Hillsdale, NJ: Lawrence Erlbaum.

——, and A. Ortony. 1977. The representation of knowledge in memory. In *Schooling and the Acquisition of Knowledge*, edited by R. C. Anderson, R. J. Spiro, and W. E. Montague, 99–135. Hillsdale, NJ: Lawrence Erlbaum.

Rundus, D. 1973. Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior* 12 (February): 43–50.

Smith, E. E. 1978. Theories of semantic memory. In *Handbook of Learning and Cognitive Processes*, edited by W. K. Estes, 1–56. Hillsdale, NJ: Lawrence Erlbaum.

Weber, R. 1980. Some characteristics of the free recall of computer controls by EDP auditors. *Journal of Accounting Research* 18 (Spring): 214–41.

Winer, B. J. 1971. *Statistical Principles in Experimental Design.* 2d ed. New York: McGraw-Hill.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 259–276

Auditors' Evaluation of Test-of-Control Strength

Eric E. Spires
The Ohio State University

SYNOPSIS: Evaluating the strength of audit evidence is a primary activity of auditors, but little is known of the factors that affect these evaluations. This article identifies and investigates two factors that affect auditors' evaluations of evidence strength.

The first factor deals with underlying constructs of evidence, called strength dimensions in this study. Although the auditing literature suggests that evidence strength is determined by such underlying constructs (see, e.g., AAA ASOBAC 1973; AICPA 1980, 1988), prior research has focused on the surface characteristics of evidence. Hence, "source of evidence" has been studied (see, e.g., Joyce and Biddle 1981; Knechel and Messier 1990; Rebele et al. 1988) rather than the underlying strength dimension of "validity" (see AICPA 1980, AU 326). The present study focuses on the underlying dimensions to determine whether they provide a basis for understanding auditors' evaluations of evidence strength.

The second factor concerns accounting firms' policies for auditing procedures. Firms' policies represent different methods of conducting efficient and effective audits (Cushing and Loebbecke 1986), and it is likely that these policies influence field auditors' judgments. Therefore, prior research on how accounting firms' degrees of structure affect auditors' judgments is extended to include the effects of specific policies.

Auditors' evaluations of test-of-control strength were used to assess whether underlying dimensions of evidence strength and the specific policies of accounting firms affect auditors' judgments. Tests of controls (TCs) are auditing procedures that determine the effectiveness of an internal control; TC strength is the degree of assurance about a control's effectiveness provided by evidence generated by a TC. In this study, TC strength is de-

This article is partly based on my dissertation, which was completed at the University of Illinois at Urbana-Champaign. I greatly appreciate the comments and suggestions of my committee members, P. Laughlin, J. McKeown, F. Neumann, and J. Wardrop, and especially those of the chairman, I. Solomon. J. Dillard, R. Libby, J. Mutchler, J. Yardley, a referee, and an associate editor. Any errors or omissions are, of course, my responsibility. The financial support of the Deloitte, Haskins & Sells Foundation is gratefully acknowledged.

Submitted July 1988.
Revisions received April and November 1989, July and October 1990.
Accepted October 1990.

fined and evaluated along the dimensions of validity, verifiability, and coverage.

Auditors evaluated five TCs along the three strength dimensions to provide dimension-based rankings of TC strength. Other auditors participated in a laboratory experiment that involved control risk assessment. Correlations between the implied ranking of TC strength from the experiment and the dimension-based rankings are high, positive, and statistically significant, which suggests that auditors' judgments of evidence strength reflect the professional standards. Conversely, studies of audit adjustments have shown that evidence strength (measured as the ability to initially detect errors) does not strongly reflect the standards. The discrepant results suggest that characterizations of evidence strength may need to be studied further.

To assess the effect of accounting firm policies on judgments, I develop predictions about the effects of policies that require reperformance and those that caution about the relative weakness of inquiry and observation. Because the predictions are largely confirmed by auditors' judgments in the laboratory experiment, accounting firms' policies are deemed to influence audit judgments. If generalizable, this finding suggests that results of research studies may be affected by the policies of accounting firms supplying subjects or data, to the extent that firms' policies differ. Researchers should therefore become familiar with firm-specific policies. For practitioners, this finding indicates that setting firm-wide policies may be effective in influencing field auditors' judgments.

Key Words: *Audit evidence, Tests of controls, Accounting firm effects, Internal control.*

Data Availability: *Data will be made available upon completion of a related project and can be obtained from the author at that time.*

THE remainder of this paper is organized as follows. Tests of controls are discussed in section I. In section II, the hypotheses are developed. The method and results are described in section III and section IV, respectively. In section V, the conclusions and research implications are presented.

I. Tests of Controls

Tests of controls (TCs) are auditing procedures that assess the operating effectiveness of an internal control structure policy or procedure (AICPA 1988, pars. 34–5).[1] Five TCs used to gather evidence about the effective operation of a control are:[2]

---

[1] SAS No. 55 (AICPA 1988, par. 34) states that TCs are directed toward assessing the effectiveness of either the *design* or *operation* of an internal control structure policy or procedure. The present study pertains only to TCs directed toward assessing operating effectiveness.

[2] SAS No. 55 specifically mentions D, I, O, and R; Libby et al. (1985) discuss S. A comprehensive literature review and discussion with auditors yielded no other TCs.

Figure 1
Summary of Control Objectives and Tests of Controls

Objective	The control has been performed independently		
Evidence	Independent employee claims, in writing, to have performed the control	Independent employee orally claims to have performed the control	Independent employee appears to have performed the control
Tests of controls	Document inspection	Inquiry	Observation

Objective	The control has been performed correctly			
Evidence	Employee implicitly claims to have performed the control correctly	Employee orally claims to have performed the control correctly	Employee appears to have performed the control correctly	Document on which the control was performed is free from error
Tests of controls	Document inspection	Inquiry	Observation	Reperformance, scanning

1. *Document inspection* (D)—inspection of documents for indication that a control has been performed by particular employees of the auditee.
2. *Inquiry* (I)—discussion with the auditee's employees about the performance of a control.
3. *Observation* (O)—observation of the performance of a control.
4. *Reperformance* (R)—performance by the auditor of the same control performed by the auditee's employees.
5. *Scanning* (S)—reviews of documents by the auditor for obvious errors that relate to a control.

The procedures can be, and usually are, combined.

To assess the effectiveness of a control, an auditor must obtain evidence that the control has been performed independently and correctly (AICPA 1988, par. 35). The first objective (independence) involves identification of the auditee's employee(s) who performed the control. The second (correctness) pertains to whether the control has been performed without error and in a timely manner.[3] Each TC can produce evidence about one or both objectives. Figure 1 presents the two control objectives, the evidence that can be obtained about the objectives, and the TCs that produce the evidence.

---

[3] A third control objective, that the control has been performed consistently, is included in SAS No. 55 (par. 35), but is excluded from the present study because it is not conceptually independent of the other control objectives. (Conceptual independence of the constructs is required so that TCs can be evaluated along the constructs without counting a particular TC's benefits and detriments more than once.) "Consistently performed" deals with whether the control was performed independently and correctly (the other two objectives) over the entire audit period and, therefore, is considered part of the "coverage" strength dimension.

TC Strength

TC strength refers to the degree of assurance (about a control's operating effectiveness) provided by evidence generated by a TC. Three dimensions of TC strength are identified in the literature:[4]

1. *Validity*—the tendency of evidence to truly represent the state of nature, where the state of nature is one of the control objectives. This dimension pertains to the degree of correspondence between a control objective and evidence produced by a TC (see figure 1).
2. *Verifiability*—the likelihood that two or more auditors, using the same TC on the same sample and at the same time, would draw the same conclusion about what was perceived. This dimension is similar to the common audit concept of "objectivity."
3. *Coverage*—the feasibility with which the entire audit period (or interim period, if applicable) can be tested using the TC.

These strength dimensions represent underlying constructs of audit evidence and can be viewed as explaining auditors' interest in the surface characteristics routinely considered in an audit. For example, the "source of evidential matter" is considered by auditors because it influences the validity of the evidence.

For purposes of this study, the dimensions used to define TC strength should be conceptually independent (see footnote 3). Clearly, coverage is conceptually independent with respect to both validity and verifiability. The conceptual independence of validity with respect to verifiability is illustrated by consideration of document inspection as it relates to the control objective of independence. Validity deals with what the presence of, say, a clerk's initials on documents indicates about whether a control was performed independently. In contrast, verifiability deals with whether auditors would agree about the presence of initials on documents. Verifiability does not deal with whether auditors agree about the meaning of the presence of initials; instead, this relates to the validity of the evidence.

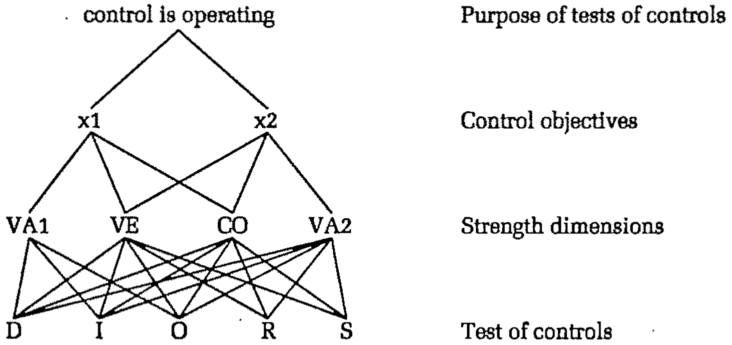
For appropriate evaluation of TC strength, the relationships between the strength dimensions and control objectives must be specified. Of the three dimensions, only validity is assumed to depend on the control objectives. That is, the validity of evidence produced by a TC may differ for each objective. Conversely, the verifiability of the evidence and the coverage provided by the evidence do not depend on the objectives and are presumed to be the same for a TC for both objectives.

The relationships among the strength dimensions, control objectives, and TCs are summarized in figure 2. As the figure shows, to determine if a control is operating effectively, the auditor needs assurance that the control was performed independently and correctly. To evaluate the quality of assurance obtained about the control objectives, the auditor considers the strength of evidence produced by TCs. This strength is defined in terms of validity, verifiability, and coverage. Although the framework presented in figure 2 is only one possible representation of TC strength, specifying the relationships among the variables in a hierarchical fashion enables one to derive numerical representations of relative TC strength, as presented in section III.

---

[4] These three strength dimensions are intended to effectively summarize the many constructs of evidence mentioned in the literature (see, e.g., AAA ASOBAC 1973; AICPA 1980, 1988; Mautz and Sharaf 1961; Schandl 1978) and thus do not map perfectly to any one source.

Figure 2
Framework for Test-of-Control Strength



x1 = control performed independently
x2 = control performed correctly
VA1 = validity for x1
VA2 = validity for x2
VE = verifiability
CO = coverage

D = document inspection
I = inquiry
O = observation
R = reperformance
S = scanning

II. Hypotheses Development

One purpose of this study is to investigate whether auditors' judgments of TC strength reflect the underlying strength dimensions. Because auditors are trained in evidence evaluation and because the professional standards discuss surface characteristics that are manifestations of the strength dimensions, it is hypothesized that judgments are consistent with the dimensions. If this consistency exists, rankings of TC strength on the basis of the strength dimensions should be positively associated with rankings that are based on auditors' judgments in a common auditing task.[5]

H1: The ranking of TC strength implied by auditors' judgments is positively correlated with dimension-based rankings of TC strength.

The other purpose of this study is to assess the influence of accounting firms' specific policies on auditors' judgments. Accounting firm effects, if they exist, would likely influence judgments about the magnitudes between ranks of TC strength, not judgments about the ordinal rankings. Ordinal rankings are based on evidence constructs contained in professional standards and should be comparable across firms. Accounting firm policies, however, fine-tune the standards and, accordingly, are likely to affect auditors' perceptions of the magnitudes between ranks. Therefore, the remaining hypotheses, which deal with effects of accounting firm policies, are stated in terms of magnitudes between ranks.

Two differences in accounting firm policies identified by Cushing and Loebbecke (1986) are used to develop expectations about differences in magnitudes between ranks. The first difference among firms relates to reperformance's role in testing controls

---

[5] So as not to induce the association experimentally, one group of auditors provides the dimension-based rankings and another group provides the judgments of TC strength.

(Cushing and Loebbecke 1986, 24; Sullivan 1984). Some firms require reperformance for certain (or all) controls, whereas other firms allow the auditor to decide whether to use reperformance. Auditors from firms requiring reperformance would tend to view it as more important than would auditors from other firms. Therefore, it is hypothesized that auditors from firms requiring reperformance derive more assurance from using reperformance than do auditors from other firms.

H2.1: Auditors from firms requiring reperformance derive more assurance from its use than do other auditors.

The second difference among firms deals with the role of inquiry and observation in testing controls that provide documentary evidence of performance (Cushing and Loebbecke 1986, 25). Some firms caution that inquiry and observation are weaker than other TCs. Auditors from these firms would likely derive less assurance from inquiry and observation (relative to other TCs) than would auditors from firms that do not provide such caution.

H2.2: Auditors from firms providing caution that inquiry and observation are relatively weak procedures derive less assurance from using inquiry and observation than do other auditors.

III. Method

Testing the hypotheses requires both dimension-based rankings and auditors' judgments of TC strength. The development of dimension-based rankings is presented in the first major subsection. The second major subsection describes the laboratory experiment used to elicit auditors' judgments of TC strength.

Dimension-Based Rankings of TC Strength

Data Collection. The analytic hierarchy process (Saaty 1980) was used to develop dimension-based rankings of TC strength. The required ratings were gathered by means of a 14-page questionnaire administered to ten auditors in the presence of a researcher. Because seniors and supervisors usually evaluate internal control and results of TCs (Abdolmohammadi and Wright 1987, 6; Haskins 1987, 557), accounting firms were asked to provide auditors who had at least senior-level experience. Participating auditors represented six accounting firms and had, on average, more than three years of auditing experience. Only one auditor had fewer than 27 months of auditing experience. All respondents had experience with TCs.

The questionnaire included definitions and discussions of the control objectives, the strength dimensions, and the TCs and their overall purpose. A nonauditing example of how to make the various ratings was also provided. For each rating, a respondent would first rank the elements (either TCs or strength dimensions) in order of strength (if ranking TCs) or importance (if ranking strength dimensions). For example, if a respondent were rating the five TCs along the verifiability dimension, the TCs would be ranked in order from the one providing the most verifiable evidence to the one providing the least. Ties were permitted. Each respondent would then compare elements in adjacent ranks[6] and assign numerical values that represented the degree of difference between the elements. For example, a value of "three" in rating verifiability would indicate that the higher-ranked TC is thought to provide evidence three times as verifiable as that provided by the lower-ranked TC. In the analysis, these rankings are converted into "priority" vectors whose elements sum to one and represent the relative importance or strength of the items being rated.

---

[6] This method of making comparisons differs from that suggested by Saaty (1980), but has been used by others (e.g., Klee 1971). Saaty's method required up to 90 minutes for some pilot-test subjects. Because partici-

Four sets of comparisons were elicited:

1. The importance of the strength dimensions in assessing the evidence produced by TCs.
2. The strength of the five TCs along each dimension. This includes four separate ratings because TCs were rated along: (1) validity for the objective of independence, (2) validity for the objective of correctness, (3) verifiability, and (4) coverage, as depicted in figure 2.
3. The strength of the five TCs in providing evidence for the achievement of the overall purpose of TCs (i.e., to determine whether a control is operating effectively).
4. The strength of four combinations of TCs (i.e., DR, DS, IOS, IO)[7] and one single TC (D) in providing evidence for the achievement of the overall purpose of TCs. The combinations chosen are commonly used in practice.[8]

Rankings. Panel A of table 1 presents rankings of mean strength for single TCs. The first two rankings reflect the relative strength of the TCs in determining attainment of the objectives of independence and correctness. These rankings involved aggregating, with the analytic hierarchy process, respondents' ratings of the importance of the strength dimensions and the ratings of the TCs along each dimension (comparison sets 1 and 2 described above).[9] The third ranking represents the relative strength of single TCs for attaining the overall purpose of TCs (to determine whether a control is oper-

---

pating accounting firms agreed to provide only one hour, the comparison method was modified, which reduced the required comparisons by about half. The cost of the modified approach is that a consistency ratio, which measures internal consistency of respondents' comparisons, cannot be calculated. The worth of the consistency ratio, however, has been questioned and it is not clear what should be done if the responses are inconsistent (see Jensen 1984). Further, responses of the majority of auditors who have used the analytic hierarchy process in other studies have been internally consistent (see Arrington et al. 1984; Messier and Schneider 1988).

[7] Note that both R and S are used with D, but not with each other. This is because S can be viewed as a weak form of R, but not as a weak form of D. The distinguishing feature of S is not the speed with which the test is performed, but rather the level of scrutiny that is used. In this respect, it is possible to reperform only part of the control procedure (i.e., consider S as a weak form of R). However, when inspecting documents for indication that a control has been performed, the auditor either finds this indication or not. S cannot be considered as a weak form of D (but see footnote 11).

[8] DR, IOS, and IO were used by Libby et al. (1985). The common use of these combinations in practice was confirmed by auditors. DS and D (by itself) were added on the basis of consultations with auditors and examination of various audit manuals and programs.

[9] Comparison sets 1 and 2 were aggregated as follows. Define $\mathbf{w}$ as the three-element priority vector for the strength dimensions and $\mathbf{v}_i$ ($i=1$ to 3) as the five-element priority vectors for the TCs related to each of the three strength dimensions. $\mathbf{V}$ is the 5×3 "performance" matrix, whose column i is the vector $\mathbf{v}_i$. For each control objective, the product $\mathbf{Vw}$ yields the five-element priority vector representing the relative strength of the TCs for attaining the control objective, as presented in panel A of table 1. The theory underlying the computations can be found in Saaty (1980).

For the vectors related to independence and correctness (panel A of table 1), means were computed after calculating $\mathbf{Vw}$ for each respondent. Computing means of $\mathbf{w}$ and $\mathbf{v}_i$ first and then calculating $\mathbf{Vw}$ does not change the results appreciably.

Table 1

Rankings of Test-of-Control (TC) Strength from Questionnaire Responses

Panel A. Strength Rankings for Single TCs:

	Control Performed Independently	*Control Performed Correctly*	*Overall Purpose*
Document inspection (D)	0.54 (1)	0.34 (1)	0.49 (1)
Inquiry (I)	0.19 (3)	0.15 (3)	0.17 (2T)
Observation (O)	0.22 (2)	0.12 (4)	0.17 (2T)
Reperformance (R)	0.00 (5)	0.29 (2)	0.10 (4)
Scanning (S)	0.05 (4)	0.10 (5)	0.07 (5)
Kendall's W	0.71	0.13	0.45

Panel B. Strength Rankings for TC Combinations:

	Elicited	*Additive*
Document inspection and reperformance (DR)	0.45 (1)	0.25 (1)
Document inspection and scanning (DS)	0.22 (2)	0.23 (2)
Document inspection only (D)	0.10 (4)	0.21 (3)
Inquiry, observation, and scanning (IOS)	0.15 (3)	0.17 (4)
Inquiry and observation (IO)	0.08 (5)	0.14 (5)
Kendall's W	0.59	

Note: Entries are means of responses. Ranks are presented in parentheses.

ating effectively). Because respondents rated single TCs directly on the overall purpose (comparison set 3 above), no aggregation is required for this rating.

Table 1 also presents values for Kendall's W, a nonparametric measure of concordance that can range from 0 (no concordance) to +1 (perfect concordance) (Siegel 1956, 229–38), to assess the similarity of the respondents' ordinal rankings. For the rankings concerning independence ($W=0.71$) and overall purpose ($W=0.45$), W is significantly greater than zero ($p<0.001$), which indicates that the respondents' rankings are similar. For the rankings concerning correctness, however, the respondents' rankings are not similar ($W=0.13$, $p=0.27$).[10] The rank correlation between the rankings concerning independence and correctness is 0.30 ($p=0.32$), which suggests that respondents perceived TCs as producing evidence of different relative strengths for the two control objectives.[11]

---

[10] It is not possible to specify exactly why consensus is lower for the objective of correctness than for the other variables. However, detailed analysis of the individual respondents' rankings for this objective suggests that respondents may have been emphasizing different components of the objective. "Performing the control correctly" implies that the control *was* performed, that it was performed in a timely manner, and that it was performed without error (i.e., the mechanics were correct). It is reasonable that D would be considered most useful in determining that the control was performed, I and O would relate well to "in a timely manner," and R and S would be considered best for judging the mechanics of the performance. Emphasis on different components of the objective may have caused low consensus. Unfortunately, most firms' audit manuals are not detailed enough to assess whether differences in firm policies could be contributing to the low consensus.

[11] The definition of S precludes it from providing evidence relative to the objective of independence, but some respondents did not recognize this (S has a nonzero weight in the ranking for the objective of independence).

Figure 3
Representation of the Experimental Design

		b_1	b_2	b_3			b_1	b_2	b_3
a_1	s_1	c_1	c_2	c_3	. . . a_4	s_{31}	c_1	c_2	c_3
	s_2	c_2	c_1	c_5		s_{32}	c_2	c_1	c_5
	s_3	c_3	c_4	c_2		s_{33}	c_3	c_4	c_2
	s_4	c_4	c_5	c_1		s_{34}	c_4	c_5	c_1
	s_5	c_5	c_3	c_4		s_{35}	c_5	c_3	c_4
	s_6	c_1	c_2	c_4		s_{36}	c_1	c_2	c_4
	s_7	c_2	c_3	c_5		s_{37}	c_2	c_3	c_5
	s_8	c_3	c_4	c_1		s_{38}	c_3	c_4	c_1
	s_9	c_4	c_5	c_2		s_{39}	c_4	c_5	c_2
	s_{10}	c_5	c_1	c_3		s_{40}	c_5	c_1	c_3

a = public accounting firm (four levels);
b = type of error that could occur in financial statements (three levels), see figure 4;
c = test-of-control combination (five levels), see figure 5; and
s = individual auditor-subject (ten from each accounting firm).

Because auditors often combine TCs, rankings of TC combinations were developed and are presented in panel B of table 1. The first ranking represents the relative strength of TC combinations for the overall purpose of TCs and was elicited directly from the respondents (comparison set 4). The auditors' responses displayed moderately high consensus on the elicited combination rating ($W = 0.59$, $p < 0.001$).[12]

The second ranking of TC combinations was developed by adding appropriate weights from the panel A ranking on the overall purpose. For example, IOS in the additive ranking is based on the sum of 0.17, 0.17, and 0.07 from the panel A ranking. After sums were computed for the five TC combinations, they were normalized to sum to one, in accordance with the analytic hierarchy process. The ordinal rankings of the additive and elicited vectors are the same except for D and IOS.

The rankings of TC combinations in panel B are based on auditors' evaluations of the strength dimensions; that is, they are based on concepts of evidence strength reported in the literature. If these rankings are similar to the judgments of TC strength auditors made in a common auditing task, the strength dimensions may provide a basis for explaining the judgments. Because D and IOS are ranked differently in the elicited and additive rankings, both rankings are compared to auditors' judgments of TC strength.

Method for Eliciting Judgments of TC Strength

Experimental Design. I used a repeated-measures incomplete Latin square design (Cochran and Cox 1957, chap. 13; Winer 1971, chap. 9), depicted in figure 3, to elicit judgments. Each subject in this design is required to make three different control assessments for a single auditee. Each assessment relates to a different type of financial statement error (and a different set of controls). Auditee-specific data do not change across assessments.

---

[12] Kendall's W of 0.59 for the elicited ranking of TC combinations includes the effect of one respondent whose ranking was negatively correlated with all other respondents' rankings. If this respondent's ranking is deleted from the analysis, W is 0.79 ($p < 0.001$). Although the ordinal ranking in table 1 does not change, the difference between D and IOS is reduced from 0.05 to 0.01, and IO is rated much lower.

The design used here provides for manipulation of variables other than TC strength, which should reduce the possibility of subjects' discerning the research purposes. Further, because treatments are assigned according to a Latin square, *two* variables change concurrently for each case within subjects. This makes it difficult for subjects to develop a simple response set; that is, subjects are less likely to condition responses for one case on responses to previous cases.

The Experimental Task. Subjects were presented first with background information about a fictitious auditee. Included was information about the nature of the auditee's business, the size and financial condition of the company, the integrity of management, and a detailed description of internal control over sales, including identification of four types of error that could occur. A sample case pertaining to the first type of error was presented to acquaint subjects with the nature of the required responses.[13]

After the background information and sample case, three "actual" cases, each pertaining to one type of financial statement error, were presented. Each case described the type of error that could occur and identified three auditee controls related to the error. On the basis of this information, subjects first evaluated the overall control strength (with the assumption that controls were operating). An audit program containing planned TCs and anticipated test results was presented. Each subject then assessed the effectiveness of the controls on the basis of the strength of controls, planned TCs, and anticipated test results. After the cases were presented, subjects completed a short debriefing questionnaire. Two pilot tests to assess the experimental materials used both auditors and academics with auditing experience.

The Independent Variables. The design in figure 3 includes four independent variables. The TC variable, which is of primary interest, had the same five levels as in panel B of table 1: DR, DS, D, IOS, and IO. Libby et al. (1985), who conducted the only other study of TC strength, used three levels (DR, IOS, and IO) and tested only DR versus IO and IOS versus IO. Additional levels were used in the present study to provide more comprehensive tests and because of their common use in practice.

Figure 4 presents the three types of financial statement error and the related controls (three for each error). The TC was manipulated (between subjects) for only one control per error. These manipulations are presented in figure 5. The other two controls always had the same TC. Manipulating TCs for only one control provides for a conservative test of TC strength because: (1) it should not have been obvious to subjects what was being manipulated, and (2) the control assessments involve combining all three controls. Therefore, the effect of testing any one control would be detectable only if it made a relatively large difference. Yet, including only three controls per error reduced the risk that subjects did not attend to the variable of primary importance (i.e., the TCs).

The Dependent Variable. Subjects were asked to assess the amount of reliance they would place on the controls (RELIANCE) on a nine-point scale. The scale descriptions

---

[13] Two sets of responses, both reasonable and accompanied by adequate explanations, were presented for the sample case. Including different sets of responses emphasized the "professional judgment" required in the cases and reduced the possibility that subjects would become predisposed to a particular response pattern.

Figure 4
Levels for the Type-of-Error Independent Variable

ERROR 1: Goods are shipped, but not billed.

Controls related to type of error b_1:

1. There is adequate segregation of functions among data processing, billing, shipping, and inventory.
2. There is adequate control over outgoing shipments.
3. Shipping documents are accounted for.*

ERROR 2: Sales are billed, but not recorded.

Controls related to type of error b_2:

1. The completeness of the sales entries is verified in a timely manner.*
2. The accounts receivable ledger is independently reconciled with the general ledger control account.
3. There is adequate segregation of functions among data processing, billing, and accounting.

ERROR 3: Pricing and/or math errors are made on sales invoices.

Controls related to type of error b_3:

1. Standard prices and discounts are used.
2. There is an independent review of the accuracy of the sales invoices, including a review of the prices and discounts used.*
3. Customer complaints are independently investigated.

* The method of testing was manipulated for controls designated by an asterisk (see figure 5).

of "no," "little," "moderate," and "maximum" reliance, as in Libby et al. (1985), were employed to maintain a strong tie to the four-point scale commonly used in practice (Elliott 1983, 8; Gafford and Carmichael 1984, 138).

The Covariate. The dependent variable RELIANCE is a function of subjects' perceptions of inherent risk and control risk. Control risk, in turn, is a function of control strength, TC strength, and TC results (Libby et al. 1985, 215). Since the experiment is used to examine the effects of TC strength on assessments of control risk, it is desirable to control for inherent risk, control strength, and TC results. TC results were controlled as always being positive. That is, there were no deviations or exceptions for any TCs. Subjects' perceptions of inherent risk and control strength were controlled through use of a covariate. For each type of error (figure 4), subjects used a nine-point scale to assess the strength of the control system. These assessments capture perceptions of both inherent risk and control strength and, therefore, are used as covariates in the analysis.

Administration of the Experiment. Forty auditors (ten from each of four accounting firms) participated in this experiment. Their auditing experience ranged from 15 to 90 months, with an average of 42 months. Only five subjects had fewer than two years' experience.

Each firm's auditors were randomly assigned to rows of the design (see figure 3). The treatment order within each row was randomized. Subjects completed the experiment in the presence of a researcher, which provided assurance that no subjects collaborated. The experiment was completed in 45 to 60 minutes.

Figure 5
Levels for the Test-of-Control (TC Strength) Independent Variable

---

Levels for Type of Error b_1 (Shipping-Document Control):

D: Randomly select 59 shipping documents from the sequence used throughout the audit period. (The sample size is based on zero expected deviations, a five percent risk of overreliance, and a five percent tolerable deviation rate. If deviations are discovered, either the sample size will be increased or the amount of control reliance will be reduced.) Obtain the shipping documents and examine each for the initials of a billing clerk (indicating that the shipping document was matched with its sales invoice, the customer signature was noted, and the shipping and inventory clerks' initials were noted).

R: For the 59 shipping documents selected above, obtain the related sales invoices. Compare each shipping document, sales order form, and sales invoice, noting agreement of (a) the customer name, number, and address, and (b) quantities and descriptions of goods sold. On the shipping document, note the presence of (a) the customer signature, and (b) the initials of the shipping and inventory clerks.

S: For the 59 shipping documents selected above, obtain the related sales invoices. Scan the shipping documents, sales order forms, and sales invoices, noting any obvious omissions or disagreements.

IO: Inquire of billing about the procedures used in matching shipping documents, sales invoices, and sales orders. Observe the procedures being performed.

Levels for Type of Error b_2 (Completeness-of-Sales-Entries Control):

D: Obtain the weekly sales journals and "log of sales invoice sequences verified" from billing. Examine ten judgmentally selected sales journals for the initials of a billing clerk (indicating that the sales journal total agrees with the sum of the daily batch control totals). Examine the "log of sales invoice sequences verified" for the audit period, noting that it is complete and current.

R: For the ten sales journals selected above, compare the total per sales journal with the daily batch totals. Investigate any exceptions. Judgmentally select ten verified numerical sequences from the "log of sales invoice sequences verified" and verify that all sales invoices in the sequence have been recorded in the sales journal.

S: Scan the ten sales journals selected above, noting any obvious errors, unusual items, or unrecorded sales invoice numbers.

IO: Inquire of billing about procedures used to verify the accuracy and completeness of sales entries and to account for the numerical sequence of sales invoices. Observe the procedures being performed.

Levels for Type of Error b_3 (Review-of-Accuracy Control):

D: Randomly select 59 sales invoices from the sequence used throughout the audit period. (The sample size is based on zero expected deviations, a five percent risk of overreliance, and a five percent tolerable deviation rate. If deviations are discovered, either the sample size will be increased or the amount of control reliance will be reduced.) Obtain the invoices and examine each for the initials of a billing clerk (indicating that prices were verified and mathematical accuracy was reviewed).

R: Obtain the standard price lists. For the 59 invoices used above, verify that the prices used were proper. Verify the mathematical accuracy of the invoices.

S: Scan the 59 invoices obtained above, noting any obvious pricing or math errors.

IO: Inquire of billing about the procedures for verifying prices and reviewing the mathematical accuracy of sales invoices. Observe the procedures being performed.

For TC combinations, all procedures in the combination would be listed, with minor rewordings as appropriate.

Treatment levels in the experiment: DR, DS, D, IOS, IO.

Table 2
ANCOVA Table

Source	SS	df	MS	F	Probability
Between Subjects					
Firm	16.34	3	5.45	1.57	0.224
Subjects within firm	124.45	36	3.46	2.23	0.017
Within Subjects					
Type of error	0.38	2	0.19	0.12	>0.25
TC strength*	27.72	4	6.93	4.48	0.007
Firm × type of error	10.00	6	1.67	1.08	>0.25
Firm × TC strength	15.04	12	1.25	0.81	>0.25
Type of error × TC strength**	13.65	7	1.95	1.26	>0.25
Firm × type of error × TC strength**	35.67	21	1.70	1.10	>0.25
Covariate	8.55	1	8.55	5.53	0.026
Residual	41.75	27	1.55		
	293.55	119			

* Sum of squares is adjusted for blocks (Cochran and Cox 1957, 510–11).

** The relative within-subject information available for these interactions is 7/8 (Winer 1971, 718–30).

IV. Results

The significance of the TC strength variable ($p=0.007$) in the analysis of covariance (table 2) indicates that subjects attended to the TC strength manipulation.[14] The nature of the significant main effect is indicated in the "overall" column of table 3, which presents mean RELIANCE ratings. *A posteriori* tests using Duncan's new multiple range test (Kirk 1982) indicate three different classes of perceived TC strength. DR and DS means are significantly larger than both the IO mean ($p<0.01$) and the D and IOS means ($p<0.05$). Further, D and IOS means are significantly larger than the IO mean ($p<0.05$). All other comparisons are nonsignificant. Thus, DR and DS were perceived to be the strongest TC combinations, followed by D and IOS. IO was perceived to be the weakest combination. These results parallel those of Libby et al. (1985), who reported that both DR and IOS were perceived to be stronger than IO.

Hypothesis H1. The study's first hypothesis states that the ranking of TC strength implied by auditors' experimental judgments is positively correlated with dimension-based rankings. The Spearman rank correlation between the table 3 "overall" ranking and the "elicited" ("additive") dimension-based ranking in panel B of table 1 is 0.87 (0.97).[15] Both correlations are significant at the 0.05 level.[16] The rankings, however, are

---

[14] The significance levels in table 2 assume that the circularity assumption for within-subjects tests is satisfied (see Kirk 1982, sec. 6.4; Winer 1971, sec. 9.7). With the conservative test suggested by Winer (1971, 712), which assumes no circularity, TC strength is still significant at traditional levels.

[15] The correlations treat D and IOS in the experiment as different ranks, even though they are very close in magnitude. If D and IOS are considered to be tied, the Spearman rank correlation with the elicited (additive) dimension-based ranking is 0.95 (0.95), with $p<0.05$.

[16] Note that tests of consistency between dimension-based and experiment-based rankings are based on means, instead of individual scores. Because means suppress some of the variance across subjects, it is unclear what conclusions can be drawn about individual auditors. Although analysis based on individuals could be undertaken for the dimension-based rankings, it could not be performed for the experiment because subjects in the experiment received only three of five possible TC strength levels (and each pertained to a different control).

Table 3
Treatment Means and Ranks

		Accounting Firms			
	Overall	*1*	*2*	*3*	*4*
DR	6.2 (1T)	6.0 (3)	6.2 (1T)	6.5 (1)	5.9 (2)
DS	6.2 (1T)	6.8 (1)	6.2 (1T)	5.5 (3)	6.3 (1)
D	5.5 (3)	5.9 (4)	5.4 (4)	5.0 (4)	5.7 (3)
IOS	5.4 (4)	6.4 (2)	5.5 (3)	5.6 (2)	4.0 (4)
IO	4.4 (5)	5.5 (5)	4.1 (5)	4.4 (5)	3.7 (5)
Mean	5.5	6.1	5.5	5.4	5.1
DR minus D (hypothesis H2.1)		0.1	0.8	1.5	0.2
Highest minus IO (hypothesis H2.2)		1.3	2.1	2.1	2.6

Note: Entries are mean responses. Ranks are presented in parentheses.
DR = combination of document inspection and reperformance;
DS = combination of document inspection and scanning;
D = document inspection;
IOS = combination of inquiry, observation, and scanning; and
IO = combination of inquiry and observation.

not exactly the same. In particular, DR is ranked more highly than DS in both dimension-based rankings, but DR and DS are tied in the experimental ranking. Also, D is ranked more highly than IOS in the additive dimension-based ranking and experimental ranking, but the opposite occurs in the elicited dimension-based ranking. Otherwise, the rankings are the same. Thus, auditors' TC strength rankings in the experimental task are largely consistent with rankings based on the evidence dimensions, which supports hypothesis H1.

Hypotheses H2.1 and H2.2. In developing hypotheses H2.1 and H2.2, I suggested that ordinal rankings of TC strength would be similar across firms. Kendall's *W* for the four firm rankings in table 3 is 0.79, significant at $p<0.05$, which indicates that the rankings are similar. However, the rankings are not equivalent. In general, differences across the firm rankings (as with differences in rankings for hypothesis H1) primarily relate to rankings of DR and DS and rankings of D and IOS.

Hypotheses H2.1 and H2.2 deal with magnitudes between ranks.[17] Hypothesis H2.1 states that auditors from firms requiring reperformance derive more assurance from its use than do auditors from other firms. Firms 2 and 3 require reperformance, whereas firms 1 and 4 allow the auditor to decide whether to use reperformance. Thus, the difference between means for DR and D, which measures the assurance provided by reperformance, should be larger for firms 2 and 3 than for firms 1 and 4. The differences of 0.8 and 1.5 are larger than the differences of 0.1 and 0.2 (table 3), and the mean difference for firms 2 and 3 is significantly larger than the mean difference for firms 1 and 4 ($p<0.05$). Thus, hypothesis H2.1 is supported.[18]

---

[17] In accordance with Winer (1971, 384), specific comparisons involved in hypotheses H2.1 and H2.2 are tested for significance even though the firm-by-TC-strength interaction (table 2) is nonsignificant. Planned comparison tests are used (Winer 1971, sec. 5.17).

Hypothesis H2.2 states that auditors from firms providing caution that inquiry and observation are relatively weak procedures derive less assurance from inquiry and observation than do auditors from other firms. The relative assurance provided by IO is measured by a comparison of IO and the strongest TC combination for each firm. Firms 3 and 4 provide caution about IO, and firms 1 and 2 do not. Thus, differences between means for the strongest TC combination and IO should be larger for firms 3 and 4 than for firms 1 and 2. The difference for firm 4 (2.6) is larger than the differences for firms 1 (1.3) and 2 (2.1), but the difference for firm 3 (2.1) is larger than the difference for only firm 1 (and the same as for firm 2; see table 3). The mean difference for firms 3 and 4 is larger than the mean difference for firms 1 and 2, but only at $p<0.10$. Thus, hypothesis H2.2 is weakly supported.

V. Discussion

The research hypotheses are generally supported. Dimension-based strength rankings are highly, positively correlated with the ranking derived from auditors' judgments in the experiment. Differences in magnitudes between some ranks are related to accounting firm policy.

Dimensions of Evidence

Results indicate that underlying constructs of evidence can be used to develop rankings of TC evidence strength that are highly correlated with auditors' judgments. That is, auditors act as if they consider these strength dimensions in evaluating evidence. Because the surface characteristics of evidence commonly mentioned in professional standards are manifestations of these underlying strength dimensions, the results provide support that auditors' evaluations of evidence strength reflect the standards.

Conversely, studies of audit adjustments have shown that evidence strength (measured as the ability to initially detect errors) is not strongly consistent with how the standards characterize evidence. In particular, evidence obtained indirectly by the auditor and from sources internal to the auditee is at least as effective in initially detecting financial statement errors as evidence obtained directly from outside the auditee (Hylas and Ashton 1982, 763; Kreutzfeldt and Wallace 1986; Wright and Ashton 1989).

These discrepant results suggest that characterizations of evidence strength may need to be reconsidered and studied further. It may be that the characterization of evidence strength depends on the signal of the evidence. For example, the audit adjustment studies dealt with negative evidence, whereas the present study (and Libby et al.'s study) dealt with positive evidence. Ashton and Ashton (1988, 1990) and Knechel and Messier (1990) reported that negative and positive evidence are not perceived by auditors as symmetric. Future research might investigate whether this asymmetry extends to the relationship between evidence signals and constructs of evidence strength. In particular, although strength dimensions were reflected in auditors' judgments about positive evidence, it is not clear whether the same reflection should or would occur for negative evidence.

---

[18] The major premise underlying hypothesis H2.1 is that a policy requiring reperformance causes auditors to perceive it as a relatively stronger TC. Another plausible effect of such a policy, however, is that procedures usually performed with reperformance (e.g., document inspection) are perceived as relatively weaker when performed alone. Data in table 3 suggest that the latter may be part of the explanation for the observed differences in magnitudes between means for DR and D. Of course, both effects (reperformance perceived as stronger and document inspection perceived as weaker when performed alone) are consistent with the general hypothesis that firms' policies affect auditors' judgments.

Accounting Firm Effects

Results indicate that accounting firm policies concerning TCs affect auditors' evaluations of TC strength. If generalizable, this finding has implications for both researchers and practitioners. For researchers, results of studies may be affected by the policies of accounting firms supplying subjects or data. This implies that accounting firm effects should be controlled, a point mentioned by Kinney (1986) and others. Most recent work on accounting firm effects has concerned the firms' degrees of structure (see, e.g., Bamber and Snowball 1988; Kinney 1986; Morris and Nichols 1988), but the present study found the firms' *specific* policies to be important. Researchers should become familiar with firm-specific policies.

For practitioners, the study provides evidence that firmwide policies do affect judgments made by auditors; thus, setting policies may be effective in influencing judgments on individual audits. Of course, if this is true, firms must consider possible adverse, as well as desirable, policy effects on judgments.[19] For example, this study shows that auditors from firms providing caution that inquiry and observation are relatively weak derive less assurance from inquiry and observation than auditors from other firms. Although the present study did not address the issue, it is possible that this policy caused these auditors to overreact to any weaknesses in inquiry and observation, thus compromising audit efficiency. In contrast, the auditors from firms providing no caution may have derived too much assurance from inquiry and observation, thus compromising audit effectiveness.

The Role of Scanning

My results conform with Libby et al.'s (1985) suggestion that scanning, when used with inquiry and observation, increases reliance. In the present study, scanning also increased reliance when used with document inspection. In fact, in the common auditing task (table 3), the average auditor did not differentiate between reperformance and scanning. If auditors' judgments in the common auditing task adequately reflect the *actual* strength of scanning, scanning might be preferable to reperformance in some audit situations. The evidence produced by scanning could be nearly as strong as that produced by reperformance, but more cheaply obtained.

However, the evidence that scanning might be preferable to reperformance is inconclusive. For example, the "elicited" strength ranking (table 1, panel B) indicates that auditors perceive reperformance as stronger than scanning when used with document inspection. Also, differences may exist among auditors regarding what scanning involves and how it can be applied (see footnote 11), which could affect the strength of evidence produced by scanning. To assess the efficacy of scanning, future research might investigate: (1) whether reperformance and scanning produce evidence of similar strength, and (2) assuming scanning's strength *is* lower than reperformance's, whether auditors' judgments reflect the lower strength.

---

[19] In this respect, a potentially important finding in psychology and auditing is that variables seemingly irrelevant to a judgment (e.g., order in which evidence is processed) apparently have large, systematic effects on judgments (Ashton and Ashton 1988; Hogarth and Einhorn 1990). Thus, in setting policies, practitioners must also be aware of these types of variables to ensure that only desired effects occur.

References

Abdolmohammadi, M., and A. Wright. 1987. An examination of the effects of experience and task complexity on audit judgments. *The Accounting Review* 62 (January): 1–13.

American Accounting Association, Committee on Basic Auditing Concepts (ASOBAC). 1973. *A Statement of Basic Auditing Concepts*. Sarasota, FL: AAA.

American Institute of Certified Public Accountants (AICPA). 1980. *Statement on Auditing Standards No. 31: Evidential Matter* (AU section 326). New York: AICPA.

——. 1988. *Statement on Auditing Standards No. 55: Consideration of the Internal Control Structure in a Financial Statement Audit*. New York: AICPA.

Arrington, C. E., W. Hillison, and R. E. Jensen. 1984. An application of analytical hierarchy process to model expert judgments on analytical review procedures. *Journal of Accounting Research* 22 (Spring): 298–312.

Ashton, A., and R. Ashton. 1988. Sequential belief revision in auditing. *The Accounting Review* 63 (October): 623–41.

Ashton, R., and A. Ashton. 1990. Evidence-Responsiveness in professional judgment: Effects of positive versus negative evidence and presentation mode. *Organizational Behavior and Human Decision Processes* 46 (June): 1–19.

Bamber, E. M., and D. Snowball. 1988. An experimental study of the effects of audit structure in uncertain task environments. *The Accounting Review* 63 (July): 490–504.

Cochran, W. G., and G. M. Cox. 1957. *Experimental Designs*. 2d ed. New York: John Wiley & Sons.

Cushing, B. E., and J. K. Loebbecke. 1986. *Comparison of Audit Methodologies of Large Accounting Firms*. Sarasota, FL: AAA.

Elliott, R. K. 1983. Unique audit methods: Peat Marwick International. *Auditing: A Journal of Practice & Theory* 2 (Spring): 1–12.

Gafford, W. W., and D. R. Carmichael. 1984. Materiality, audit risk and sampling: A nuts-and-bolts approach (part two). *Journal of Accountancy* 158 (November): 125–38.

Haskins, M. E. 1987. Client control environments: An examination of auditors' perceptions. *The Accounting Review* 62 (July): 542–63.

Hogarth, R., and H. Einhorn. 1990. Order effects in belief updating: The belief-adjustment model. Working paper, University of Chicago (May).

Hylas, R., and R. Ashton. 1982. Audit detection of financial statement errors. *The Accounting Review* 57 (October): 751–65.

Jensen, R. 1984. An alternative scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* 28 (September): 317–32.

Joyce, E. J., and G. C. Biddle. 1981. Are auditors' judgments sufficiently regressive? *Journal of Accounting Research* 19 (Autumn): 323–49.

Kinney, W. R., Jr. 1986. Audit technology and preferences for auditing standards. *Journal of Accounting & Economics* 8 (March): 73–89.

Kirk, R. E. 1982. *Experimental Design*. 2d ed. Belmont, CA: Brooks/Cole Publishing Company.

Klee, A. 1971. The role of decision models in the evaluation of competing environmental health alternatives. *Management Science* 18 (October): B52–B67.

Knechel, W. R., and W. F. Messier, Jr. 1990. Sequential auditor decision making: Evidence planning and evaluation. *Contemporary Accounting Research* 6 (Spring): 386–406.

Kreutzfeldt, R., and W. Wallace. 1986. Error characteristics in audit populations: Their profile and relationship to environmental factors. *Auditing: A Journal of Practice & Theory* 6 (Fall): 20–43.

Libby, R., J. T. Artman, and J. J. Willingham. 1985. Process susceptibility, control risk, and audit planning. *The Accounting Review* 60 (April): 212–30.

Mautz, R., and H. A. Sharaf. 1961. *The Philosophy of Auditing*. Sarasota, FL: AAA.

Messier, W. F., Jr., and A. Schneider. 1988. A hierarchical approach to the external auditor's evaluation of the internal auditing function. *Contemporary Accounting Research* 4 (Spring): 337–53.
Morris, M. H., and W. D. Nichols. 1988. Consistency exceptions: Materiality judgments and audit firm structure. *The Accounting Review* 63 (April): 237–54.
Rebele, J. E., J. A. Heintz, and G. E. Briden. 1988. Independent auditor sensitivity to evidence reliability. *Auditing: A Journal of Practice & Theory* 8 (Fall): 43–52.
Saaty, T. L. 1980. *The Analytic Hierarchy Process.* New York: McGraw-Hill.
Schandl, C. 1978. *Theory of Auditing.* Houston: Scholars Book Company.
Siegel, S. 1956. *Nonparametric Statistics.* New York: McGraw-Hill.
Sullivan, J. D. 1984. The case for the unstructured audit approach. *Auditing Symposium VII* (University of Kansas): 61–68.
Winer, B. J. 1971. *Statistical Principles in Experimental Design.* New York: McGraw-Hill.
Wright, A., and R. H. Ashton. 1989. Identifying audit adjustments with attention-directing procedures. *The Accounting Review* 64 (October): 710–28.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 277–290

A Perspective on Cognitive Research in Accounting

Robin M. Hogarth
University of Chicago

THE communality between research on human cognitive processes and auditing can be characterized by two words: judgment and risk. Auditors express opinions based on investigations that, no matter how thorough, inevitably involve subjective judgments. Indeed, even with impressive, recent advances in the automation of human judgment, it seems likely that auditors will continue to rely on human judgment for many years. Risk is important because the potential costs of erroneous opinions are high.

As a scientific discipline, research on cognitive processes is still relatively young. Moreover, although it adopts many sophisticated methodological tools (e.g., controlled experiments, statistics, etc.), it labors under a major handicap. One cannot directly observe the phenomena of interest, i.e., actual cognitive processes. Instead, cognitive psychologists can develop theories only by observing inputs and outputs. The consequence is that theories about mediating processes are difficult to construct and test.[1] In fact, a brief review of the literature shows a remarkable variety of different approaches even within the branch of cognitive psychology that has explicitly studied issues of judgment and risk. For example, cognitive psychologists have used models based in economics (the expected utility hypothesis), statistics (Bayes' theorem), Brunswik's lens model (see, e.g., Slovic and Lichtenstein 1971), algebraic input-output models (Anderson 1981), and computers (Newell and Simon 1972). Moreover, this diversity has been matched by researchers interested in auditing. For example, studies have looked at auditing from the perspective of the lens model (e.g., Ashton 1974), Bayesian statistics and the literature on heuristics and biases (e.g., Joyce and Biddle 1981), analysis of variance techniques (e.g., Krogstad et al. 1984), descriptive variants of expected utility theory, learning and memory processes (Weber 1980; Libby 1985), and various aspects of process tracing methodology (see, e.g., Klersey and Mock 1989).

---

[1] This comment applies both to "process tracers" and "mathematical modelers." The former, however, do attempt to get closer to the source of the data.

I am grateful to the following for comments on this Perspective: Alison Ashton, Robert Ashton, Mike Ferguson, Diane Matson, Eric Hirst, Ed Joyce, Robert Libby, and Mark Penno.

Commissioned by the Editor.
Revised draft accepted January 1991.

Faced with the facts that underlying cognitive processes are not amenable to direct observation, and that there are many different approaches to studying cognition, one wonders how—in the light of current knowledge—researchers specifically interested in *auditing* might best orient their efforts. Because I am an observer rather than a participant in cognitive research in auditing (CRA), I am reluctant to suggest what others should do; nonetheless I shall explore this question here in the belief that an outsider might touch on issues that could be overlooked by those closer to the actual research.[2]

Cognitive psychologists who specialize in judgment and decision making typically focus their work on three questions: (1) how *well* do people perform particular judgmental tasks? (2) *how* do people perform particular judgmental tasks? (3) how can you *help* people perform better? Moreover, work is centered on specific types of judgmental tasks, such as assessments of probability, which researchers believe have applicability in a wide range of practical domains. CRA has leveraged these findings by identifying situations in auditing where limitations in human information-processing abilities might lead to judgmental biases (see, e.g., Ashton and Ashton 1988) or difficulties in making certain kinds of judgments. Therefore, although more narrowly focussed than studies in cognitive psychology, CRA is pragmatic in that the identification of mechanisms for coping with human cognitive limitations is a major research goal. Over the last two decades, CRA has amassed an extensive experimental literature and, as I shall argue below, is now well positioned to serve as a role model for applied work in cognition.

Although not all work in CRA addresses all three issues enumerated above, all three are represented by the papers in this Forum. Ashton (1991), for example, asks *how well* auditors are able to assess the base rates of errors. This, in turn, raises the issue of whether it is reasonable to model auditing within a decision-theoretic framework where auditors are expected to use base-rate information as "priors." Spires (1991) investigates whether judgments of the strength of audit evidence correspond to the "true" underlying dimensions of validity, verifiability, and coverage. This addresses the question of whether auditors' judgments reflect standards specified by the profession and is important because it questions whether auditors can perform certain tasks. Thus, if auditors' unaided judgments are deficient (in fact, this was not the case), ways should be found to remedy the deficiencies, such as by providing judgment aids. Frederick (1991) inquires how auditors encode knowledge in memory and the implications this might have concerning the construction and use of memory aids in auditing. All authors are ultimately concerned with the issue of how to help auditors make better judgments.

An interesting feature about the set of papers is that although all contain numerous references to the mainstream literature in psychology, they do not represent the same research paradigms. Frederick's paper, for example, is in the style of experimental cognitive psychology. At one level, it can be said that his subjects just happened to be auditors. Spires' paper could be described as a judgment study, almost within the lens model tradition even though he makes no appeal to that particular methodology; and Ashton's work could be thought of as a "field study" that tested for the presence of cognitive biases (Kahneman et al. 1982).[3] That two of the papers (those by Ashton and Frederick) dealt explicitly with memory processes is significant in that this underscores a growing awareness that judgments are influenced not only by data in the environment, but by data that are retrieved from memory. Thus, understanding how memory is organized (Frederick) or what data can be accessed (Ashton) are important issues. These Forum papers are not only informed by the literature in cognitive psychology, they also inform that literature.

---

[2] Although modern cognitive psychology covers many subfields, this Perspective is based largely on my familiarity with the area known as "judgment and decision making" (see, e.g., Hogarth 1987).

Because of the unobservable nature of cognitive processes, the study of judgment makes heavy use of both standards and metaphors. Standards are applied to observable outputs (i.e., actual judgments), whereas metaphors are used in the form of "as if" models of the underlying and unobservable judgmental processes. The outputs of Bayes' theorem, for example, can be used to measure the accuracy of revisions of probabilistic opinions, and the structural form of the model may also suggest a mechanism or process by which opinions are revised (cf., Hogarth and Einhorn 1990). Indeed, one of the fundamental tensions in CRA centers on the choice of metaphors and, in particular, the extent to which metaphors grounded in psychology can provide more instructive insights than those assumed within a more rationalistic, economic framework. That is, although researchers within the economic framework also postulate "as if" models or metaphors, these are used only to generate observable predictions and there is little interest in the underlying processes themselves. In CRA, however, the desire to explicate the nature of the processes drives much of the research. The rationale is simple. You cannot improve a process unless you know what it is. (For a striking practical example of this principle, see Lopes 1987.)[4]

Metaphors and standards are used explicitly in all three Forum papers. Ashton uses the metaphor of "man as a statistician" (cf., Peterson and Beach 1967) to compare human judgments of error rates with empirical relative frequencies. Further, at the end of her paper, she suggests that researchers adopt a more general metaphor—that of auditors as "general practitioners." Spires uses the standard of the Saaty (1980) method for decomposing components of judgment; and Frederick contrasts two metaphors that describe how information might be organized in memory, i.e., taxonomic versus schematic representations.

The argument I develop can be summarized as follows. No metaphor is complete. It is appropriate that CRA use different metaphors for studying specific aspects of audit judgment. However, by choosing particular metaphors, researchers make tradeoffs with respect to what their studies can reveal about the audit judgment process. My goal is to make explicit some of the tradeoffs implied by the choice of these metaphors. One important tradeoff concerns the gains achieved by studying the ability of auditors to make specific kinds of judgments as opposed to investigating their competence across the wider range of tasks with which they are confronted.

In my opinion, the challenge facing CRA is twofold. One is to continue to examine specific aspects of audit judgments. Indeed, progress depends heavily on such studies. The second and more difficult task is to elucidate the systemic properties of audits and the relative importance of cognitive processes in determining the efficiency of such systems. In other words, we also need to understand how knowledge acquired on specific aspects of judgment fits into a larger framework. To avoid misunderstanding, I emphasize that my comments are not confined to CRA. They apply more broadly to any area that imports basic knowledge from cognitive psychology to illuminate practical problems and are organized around three themes—(1) "Ought implies can," (2) How good do you need to be?" and (3) The multidimensional nature of auditing.

---

[3] I should state that I don't expect that the authors will necessarily agree with my classification.

[4] This belief stands in sharp contrast with the economic perspective that performance can be improved by appropriate incentives, which assumes that people both understand and are capable of responding to incentives.

I. Ought Implies Can

As noted above, normative standards that have been imported, mostly from microeconomics (e.g., the expected utility model) or statistics (e.g., Bayes' theorem), play an important role in CRA. They provide a benchmark against which performance is evaluated (the how well question); they suggest hypotheses concerning process (the how question); and they are invoked in the design of decision aids for auditing (the how to help question).

Although providing a useful stimulus to research, these benchmarks can extract a large price. This is the fact that researchers often overlook the notion that "ought implies can" (Goldman 1978). In other words, if a model says that a human *ought* to make a certain kind of response (e.g., retrieve an accurate error-rate from memory), the frequent assumption is that the person *can* do so. If however, the person cannot do so (for reasons to be discussed below), one must seriously question the force of the imperative ought and thus the standard itself.

There are several reasons why "oughts" may not imply "cans." Some of these reside within the individuals making judgments, and some are inherent in the nature of the tasks they are required to perform. Classic examples of the former are embodied in Simon's (1957) principle of bounded rationality. People do not have the physical information-processing capacity to perform certain kinds of reasoning or judgment tasks.

There are generally two conditions necessary for making correct judgments. One of these concerns using the appropriate *process*; the other relates to having the appropriate *inputs* (i.e., data or knowledge) on which the process acts. Human cognitive limitations clearly influence the former; however, it is different task conditions that affect the latter. From an "ought implies can" viewpoint, it is important to understand how people acquire the knowledge on which they act and whether the task conditions under which they operate permit them to acquire the appropriate knowledge. If it turns out that, through practical experience and/or formal training, people cannot acquire the knowledge necessary to make certain types of judgments, then one can no longer realistically apply the normative standard.

When considering this issue, it is illuminating to distinguish two types of situation. In one, the nature of the auditor's experience is such that he or she cannot possibly acquire the necessary knowledge. In the second, the auditor acquires knowledge through experience. However, feedback is systematically biased such that the experience gained is inappropriate.

Concerning the former, Ashton's paper nicely demonstrates that it is unreasonable to expect even experienced auditors to acquire accurate knowledge of error frequencies when such errors are relatively infrequent in audits and when auditors do not actually conduct many different audits in a year. In fact, on considering the data she presents, as well as the earlier study by Libby and Frederick (1990), I am struck by two observations. First, given that auditors do not have much experience with errors, I am surprised by the level of accuracy that Ashton's respondents achieved on some tasks.[5] In other words, if the frequencies auditors estimated were based simply on retrieving from memory what had been encoded through past experience, then their estimates should have been worse than they were! Second, given that their answers to questions about error frequencies had some validity, how were these judgments made? One possibility is that the auditors relied principally on their causal knowledge of accounting systems to generate what could be called "educated guesses."

To pursue this idea, consider the following thought experiment. Do people make more mistakes in providing answers to questions involving mathematical calculations of squares or square roots? For instance, "What is the square of 10?" as opposed to "What is the square root of 10?" Now, most of us know that the process by which one calculates the square of a two-digit number is easier to recall and implement (as a problem in mental mathematics) than the process of calculating a square root. Thus, if asked whether humans make more errors in calculating squares or square roots, I believe that most would opt for square roots (which, for the sake of argument, we shall assume is the correct response). The point of the thought experiment, however, is not which answer is correct, but that one can make intelligent guesses about the kinds of error frequencies that are more prevalent without ever having experienced the specific task. In other words, even if people do build up images of the relative frequencies of different types of events through "automatic" memorial processes (Hasher and Zacks 1979, 1984), this does not mean that all estimates involve simply retrieving those automatically encoded frequencies. Many answers could be generated by appealing to one's causal understanding of the underlying data-generating process and adopting a heuristic that assumes a simple relation between the difficulty of a task and the probability of errors.

Situations in which people—including auditors—learn the "wrong" things through experience have been discussed at some length in the literature, and I shall not elaborate on these here (see, e.g., Einhorn and Hogarth 1978; Waller and Felix 1984). However, it is important to bear in mind that, by its very nature, auditing is a task in which true experimentation is costly, so there are many things that auditors cannot learn from feedback. For example, auditors cannot leave several tasks undone in order to learn whether they are necessary for providing an audit opinion because, if an error were subsequently discovered, the audit procedures would be judged as having been inadequate.

In summary, researchers need to be aware of the "ought implies can" distinction and to use this as a guiding principle in selecting metaphors for modeling components of the audit process. In modeling cognitive processes, researchers "ought" to adopt models that people "can" implement.

II. How Good Do You Need To Be?

A closely related issue centers on how good auditors need to be at the different judgmental tasks they perform. One problem, however, with the way the "*how well*" question is framed in many studies is the lack of appreciation for the robustness of natural but suboptimal strategies for judgment across different tasks. In other words, does it matter if auditors make suboptimal judgments and, if so, how much does it matter?

---

[5] This statement is not meant to imply that the auditors in Ashton's study were accurate in estimating frequencies when their performance is compared to some tasks studied in the psychological literature. However, they seemed to be aware of the most likely errors, and a few auditors were quite accurate in certain subtasks.

If one adopts a functional, almost Darwinian approach to human judgment, it seems clear that one would not expect humans to make optimal judgments in all types of situations, including those in which they have much experience. Instead, one would expect humans to make judgments that are "good enough" in their environments, by which is meant a level of judgmental accuracy that does not result in negative feedback. Invoking such evolutionary arguments is, however, dangerous. On the one hand, the arguments can become tautological at a theoretical level thereby negating their utility. On the other hand, individual auditing firms don't want to wait for negative feedback before they know whether they have learned to make certain kinds of judgments appropriately. In the most extreme case, negative feedback means that a firm will no longer exist.

How then should CRA deal with these issues? Fortunately, we are not without clues. For example, starting with the classic work of Dawes and Corrigan (1974), we know some useful facts about the robustness of simple linear models for approximating the evaluation of complex multivariate stimuli (see also Einhorn and Hogarth 1975). In addition, several researchers have investigated how simple decision rules perform across different types of task. Thorngate (1980), for example, showed how simple rules could approximate optimal choices made by an expected utility rule, and McKenzie (1990) recently clarified much of the literature in covariation assessment by showing that certain simple cognitive strategies are remarkably valid. In a simulated medical decision-making task, Kleinmuntz (1985) further demonstrated that use of base rates might not be that important to the taking of effective actions. An important task for CRA, therefore, would be to specify different types of audit judgments and to investigate just how sensitive decisions are to errors. For researchers, there may be tradeoffs involving what can be learned from further study of certain well-structured issues as opposed to exploratory work on new topics with larger potential payoffs.

Within these kinds of investigations, however, I believe it is important to separate three kinds of questions. First, to what extent does performance of simple vis-à-vis optimal rules depend on characteristics of the rules themselves? Second, to what extent does performance depend on the knowledge possessed by the decision maker? And third, to what extent does performance depend on particular characteristics of the task?

For example, consider the judgmental task examined by Spires (1991). In his work, the main question was whether auditors' judgments of controls that were observable reflected the underlying dimensions to which auditors should have paid attention. In general, he found that this was the case. However, there is another way in which this type of question could and perhaps should also be investigated; namely, how much leeway or error is tolerable in auditors' judgments of controls? What happens if auditors don't base their judgments on the correct underlying constructs but only on related constructs? One line of research that might be relevant has considered the effects of misspecifying variables in multiattribute decision making. It turns out that degree of error depends on the extent to which the attributes people do use are correlated with the "true" attributes that they don't use. Moreover, provided there are several attributes, results indicate that choosing the right alternative is remarkably robust over a wide range of correlations (see Gibbins 1982; Kleinmuntz 1990).

These comments apply, of course, to particular judgmental tasks. However, as will be elaborated below, it is important to consider the range of judgmental tasks performed by the auditor. To what extent are different tasks correlated? To what extent can errors at one stage of the audit process be identified at other stages, and so on?

III. The Multidimensional Nature of Auditing

The Forum papers by Ashton and Frederick comment on the nature of expertise in auditing. This is discussed explicitly and followed up with what is now a familiar pattern in CRA studies, the comparison of subjects with different levels of auditing experience.

In several CRA papers, the underlying assumption for these comparisons is that expertise is correlated with length of experience as an auditor. (For a more sophisticated view and some empirical evidence, see Bonner and Lewis, in press.)[6] Thus, when researchers find no differences in results between "experts" and "novices," the claim is made that the empirical findings (couched within the *how well* question) transcend the issue of experience. This, therefore, also becomes an argument in the debate with scholars from the economic side of the auditing literature that poor performance by experimental subjects cannot be attributed to lack of either experience or expertise.

There is an alternative, and as yet untested, viewpoint that Ashton touches on at the end of her paper. This is the notion that auditors play many different roles in their jobs and thus have to develop expertise across a range of topics. It is not clear that auditors can become expert in every task they face nor (as discussed above) is it clear that they need to.

If one accepts the notion that auditing is a multifaceted task, it is important to specify both the dimensions of the task and their relative importance. Indeed, I suggest that CRA could be considerably advanced if we were to gain a better *systemic* understanding of the nature of expertise in auditing.[7] One way to think about this would be to structure a research program that asks the following kinds of questions: (1) What are the different types of judgments auditors are required to make? (2) What are the demands of these different tasks in terms of the knowledge and processes required to make the judgments? (3) How do the different tasks relate to each other? (4) Which tasks call upon audit- or accounting-specific expertise and which call upon other types of expertise? (5) Under what conditions can auditors be expected to acquire the ability necessary to perform the different tasks?

This would be an ambitious research program, but I believe it is important. By allowing research on basic issues in cognitive psychology to suggest what might be illuminating in CRA, work has been conducted in fragmented fashion. This is not to say, however, that these studies have been without interest. Indeed, many are of considerable interest to psychologists because they provide evidence concerning the extent to which findings based on one population can be generalized to another (Shanteau 1987).

It is unclear, however, whether studies conceived in this manner provide the best vehicle for promoting cumulative progress in CRA or, in fact, in any applied field. In addition, I believe that this strategy misses what I consider to be the most important general finding in modern cognitive psychology. This is that, as limited information processors, humans are adaptive. Therefore, to understand what cognitive processes humans use, it is first necessary to make a thorough analysis of the tasks with which they are confronted. This implies that when findings from cognitive psychology are used to provide hypotheses concerning auditing, researchers should specify whether these reflect basic psychological processes, on the one hand, or demands of the tasks investigated, on the other. For example, if results from studies of experts in chess were used to generate hypotheses concerning auditing, it is important to specify explicitly either why the structures of the tasks are similar or why the cognitive processes investigated can be expected to be invariant across different tasks.

---

[6] Ashton, it should be noted, does not equate expertise with experience.

[7] This comment does not, of course, apply uniquely to CRA.

To illustrate, consider findings from the literature on the psychology of expertise. On the one hand, several studies conducted within the problem-solving paradigm reveal considerable cognitive skills. For example, chess-players (see, e.g., de Groot 1965) and physicists (see, e.g., Larkin et al. 1980) have been shown to be capable of remarkable feats of judgment, reasoning, and memory. On the other hand, there is also a considerable literature that shows the poverty of expert judgment. For example, clinical psychologists are no better at predicting in some tasks than persons with no professional expertise (Oskamp 1962); and judgment in many complex tasks would be greatly improved if experts were removed and replaced by simple, linear rules (Dawes et al. 1989). The important issue, however, is not to characterize professionals in different areas as good or bad experts but to understand why expertise is easier to acquire and demonstrate in some domains than in others (see also Shanteau in press). Thus, a multidimensional view of auditing in future studies may reveal that professional auditors are quite expert in some domains, but their level of performance in other domains will not differentiate them from relative novices (cf., Ashton 1981; Bonner and Lewis in press). Nor do I believe that the latter would necessarily be a problem, either for auditors or their clients.

A further point is that it is unclear at present how expert auditors need to be in the different roles they play. How does expertise in one dimension trade off against expertise in another? How important is coordination between different areas of expertise in auditing and to what extent does the role of judgment in auditing take place within a system with redundancies that compensate for particular types of judgmental errors.

Scholars in CRA should not feel singled out by these remarks. Researchers in the field of judgment and decision making in general have a tendency to forget that specific judgments form part of a total cognitive system with which people deal with the world. Indeed, in a classic paper, Toda (1962) made the point that humans do not necessarily need to be effective at each and every cognitive task they perform; but what is important is efficiency in coordination between tasks (see also, Hogarth 1981).

Toda's point is particularly important in CRA. If, for example, one considered the total range of tasks studied in CRA (including the three Forum papers), one would conclude that auditors differ in their abilities (the *how well* question). In addition, auditors may or may not be subject to biases induced by heuristics in probabilistic reasoning (Joyce and Biddle 1981), they make systematic errors in covariation tasks (Waller and Felix 1987), they are subject to hindsight biases (Buchman 1985), their judgments reflect suboptimal order effects (Ashton and Ashton 1988), and so on. Yet, it is difficult to understand from these studies how findings relate to actual audit performance. What, for example, is the relative importance of different judgmental tasks? For which kinds of judgments are the loss functions implied by the task more or less exacting in terms of penalties for errors (cf., Hogarth et al. in press)?

Lest what I am saying be misinterpreted, I do not deny that auditors' judgments—like those of other professionals—have been shown to exhibit systematic errors and biases and that these may be important. However, we don't know the consequences of particular judgmental errors in the auditor's natural ecology. For example, how important are order effects in judgment? (See Asare 1989.) Also, how might different types of judgment errors either cancel out with others or be attenuated by redundancies in the audit environment? After all, inherent to the structure of many auditing tasks is the fact that accounting systems generate correlated observations (induced by double-entry bookkeeping). For example, higher than typical sales in the closing month of a financial year should be accompanied by higher levels of receivables or cash. Thus, to the extent that an audit misses an error in one domain, there is a chance that—in the absence of something pernicious—any discrepancy will reveal itself in another area of the accounting system.

Redundancies can, of course, also be built specifically into the way audits are conducted (e.g., second-partner reviews are used to capture a variety of possible errors). However, without a more systemic knowledge of audit judgments, it is difficult to assess whether these are the appropriate redundancies. In other words, where are the most effective locations for building redundancies into the audit system to avoid judgmental errors? Would it be more effective to build in redundancies systematically at lower, more operational levels of the audit to cover specific areas where judgments have been shown to be fallible as opposed to relying on second-partner reviews? To what extent might second-partner reviews carry the dangers of replicating error and generating false confidence?

The issue of redundancy becomes particularly important when one moves from laboratory to field studies. In the former, auditor subjects are presented with relatively isolated experimental stimuli (see, e.g., Frederick 1991) and are asked to make judgments concerning hypothetical clients. However, in the context-rich domain of real-world audit experience, it is feasible that auditors may be alerted to examining particular variables by picking up cues that would be hard to replicate in an experimental setting. For example, one domain of auditor expertise that would be difficult to study in laboratory settings is the ability to detect rare "broken-leg" cues (Meehl 1954) that alert auditors to potential problems (but see Johnson et al. 1989).

As noted, Ashton (1991) has suggested the metaphor of auditor as general practitioner, thereby explicitly recognizing the multifaceted nature of auditing and the fact that auditors may not be experts in all tasks. Indeed, in accordance with Toda (1962), the key to success in auditing may lie in the ability to coordinate a series of different tasks rather than in performing each task in an optimal manner. This is an important insight and raises the issue of how it might be investigated.

One way might be to create an experimental, computerized audit game (rather like a flight simulator) in which subjects would be required to play the role of an auditor dealing with different tasks and with the ability to consult different experts on different matters. (For examples of similar work in other fields, see Brehmer 1990; Sterman 1989. For an earlier call for this kind of work in accounting, see Swieringa and Weick 1982.) The auditor's performance would therefore not just be judged on such tasks as accuracy in specific tests of internal control but on broader criteria that would include the ability to plan the overall audit, to decide what kind and level of audit staff and specialists to use, to make tradeoffs between different types of audit tests, and so on. Further interesting dimensions could involve the effects of having to change the audit plan because of unexpected problems, the need to investigate some issues in greater depth than anticipated, and the effects of including and excluding redundancies in different components of the audit. Such games would also permit investigation of important audit management issues, such as the effects of deadlines or cost overruns with respect to time budgeted to perform specific tasks.

One implication is that, CRA studies would become less reliant on the literature in cognitive psychology and would have to make contact with other literatures in the social sciences (e.g., the literature dealing with organizational phenomena). This, in turn, suggests a need for more collaborative work with scholars from other disciplines.

Parenthetically, audit simulation games could possibly be developed in conjunction with a major accounting firm and may well yield more ecologically valid responses than the questionnaires to practitioners that currently provide data for many CRA studies. Moreover, it may be possible to integrate the use of such games within auditor training programs thereby facilitating the goals of both research and training.

A further advantage of such simulations would be to recognize that although much audit work is carried out by individuals, audit opinions are signed by firms. Thus, an important dimension of auditor expertise and judgment—particularly at the level of partner and manager—lies in more general management decisions and specifically in the coordination and motivation of teams of auditors. For example, a particularly interesting and important problem concerns how managers or partners allocate their time and that of others to different tasks or subtasks. Is this done in something approaching an optimal manner or do auditors use the matching law that Herrnstein (1990) calls the *melioration principle*? In other words, when allocating resources across a set of alternatives, do auditors allocate in a manner that equates average returns from all alternatives (i.e., matching or melioration) or does their allocation of resources reflect maximizing behavior? Alternatively, if auditors do exhibit matching behavior, can this be rationalized within the context of the auditing task?

The complexities of inference in large experimental audit simulations are, of course, daunting (cf., Ashton 1982). Thus, at a pragmatic level it is legitimate to ask how to proceed. One way could be to adopt an incremental approach. This would mean building an experimental shell that is initially used only for testing a few hypotheses. Over time, however, the shell could be made successively more complex until a fairly realistic simulation is achieved.

Alternatively, one could follow a suggestion by Brunswik (1956) that reverses the usual scientific procedure of controlling one variable at a time in the incremental fashion just described. To study whether a particular variable is important (e.g., the accuracy of a specific type of judgment), Brunswik suggested that this be done not by manipulating the variable per se, but rather by creating a system in which that variable is absent and then seeing what effect the absence of the variable has on the system (see also, Hogarth 1986). Imagine, for example, that you wanted to find out whether it was important that auditors make accurate judgments of the strength of internal controls (cf., Spires 1991). This could be examined by creating an environment where auditors cannot make accurate judgments and then observing whether this affects the overall quality of the audit. It could be the case, for example, that certain redundancies—

either in the environment or in the way audits are conducted—compensate for the lack of accurate judgment. Conversely, accuracy might be very important.

Whether full-scale or partial simulations are used, however, I believe that CRA is mature enough for more complex designs. Fortunately, some recent work points in this direction (see, e.g., McDaniel 1990).

As noted previously, CRA has been revealing in terms of what individual auditors can do in specific tasks. However, it is also important to ask whether variance in audit performance across firms (or even within firms) depends more on the quality of individual judgments or more on how teams of auditors are managed on matters that are not directly related to audit judgment *per se*. In other words, what are the different sources of variation in audit quality and which are the more important? Despite my keen professional interest in judgmental issues, I would hazard a guess that management variables are as important as, if not more important than, many judgmental issues. Both, however, need to be studied together.

An issue raised by Spires is important in this respect. Building on the work by Cushing and Loebbecke (1986), he shows that, at the level of specific audit procedures, the way auditors make judgments can depend on the nature of different practices within audit firms. It is easy to imagine that different corporate cultures within audit firms could also affect other dimensions of audit expertise and judgment (see Schein 1985). For example, the often tacit manner in which firms resolve technical disputes concerning audit policies could have significant effects on the audit opinions they issue. In addition, corporate cultures affect attitudes (e.g., toward risk and professionalism), and these may have nontrivial impacts on audit judgments.

It is important to emphasize that the above comments are not meant to belittle work that has been done on audit judgment to date. Nor am I suggesting a panacea for audit judgment studies. Instead, I am arguing for researchers to enlarge the context in which audit judgments are examined. If the key to audit judgment lies in its multifaceted nature, then this needs to be investigated.

IV. Concluding Comments

The general argument in this perspective can be restated as follows. Auditing is a multifaceted process in which judgment plays a key role. Moreover, CRA studies to date have successfully elucidated several aspects of how auditors make some of these judgments. Most of these studies, however, have focussed on particular aspects of audit judgment and have been conducted within specific metaphors of underlying processes. There have been two advantages of this approach. First, the work in CRA has become well-grounded in the discipline of psychology, both in terms of theory and methodology. CRA generally demonstrates a high level of professional competence. Second, because auditing is an important professional occupation, results have been illuminating independent of whether they do or do not corroborate earlier findings from psychology.

In the future, there will always be a need for good studies that examine specific aspects of judgmental processes in auditing. The phenomena are important and progress is often made by focussing on particular issues. However, more is needed. Because the approach taken to date typically examines tasks in isolation from other demands of the audit environment, it is difficult to understand the implications of findings, either with respect to statements concerning judgmental performance of auditors or how findings might be used to improve the practice of auditing. Also, because researchers have focussed on specific aspects of audit procedures, it is difficult to assess how important suggested changes in those parts of the process fit into the overall design of an audit. How sensitive are audits to changes made to parts of the overall process?

Research involves choices and tradeoffs. By focussing in depth on particular aspects of audit judgment, CRA has made considerable progress over the last two decades. In hindsight, a case can be made that specific, focussed studies were the appropriate way to start CRA. Nevertheless, some progress might have been gained at the expense of achieving a more systematic knowledge of auditing. To correct this imbalance, it is important to develop a more complete description of the different types of judgmental tasks auditors are required to perform. What are the different dimensions of audit judgment tasks? How do they differ in importance and under what conditions? What contextual features might have the greatest impact on audit judgment? What, for example, are the effects of time pressure and/or the riskiness of different types of audits? How do auditors make tradeoffs between different types of tests they could perform? How do auditors form plans at the level of the audit itself? How does the corporate culture of the audit firm condition the judgments of individual auditors?

These questions stress the importance of developing mechanisms for studying audit judgments within larger contexts. Although I don't expect this will be easy, one suggestion is to develop audit games that capture many of the dynamic, interdependent features of an audit within the confines of a simulated business environment. Such games should be particularly illuminating with respect to (1) the extent to which different types of judgmental errors are important or unimportant, and (2) the dimensions of auditor expertise that have most impact on overall audit quality.

Finally, many statements made here apply with equal force to areas other than auditing. Indeed, for some time I have expressed similar comments about research in the broader realm of judgment and decision making (see, e.g., Hogarth 1981, 1986). On the positive side, the work in CRA provides an important source of ideas and hypotheses concerning the types of audit judgments that are effective or ineffective, as well as the kinds of tools that may be needed to improve the audit process. In the final analysis, however, audits take place in a larger and more complex context than those portrayed in most CRA studies. It is my belief that CRA is now mature enough to consider the complexities of that context. Indeed, a new goal for CRA is to serve as a role model for how to investigate professional judgment in applied settings.

References

Anderson, N. H. 1981. *Foundations of Information Integration Theory*. New York: Academic Press.

Asare, S. 1989. The auditor's going concern decision: Interaction of task variables and the sequential processing of evidence. Unpublished manuscript, Fisher School of Accounting, University of Florida, Gainesville.

Ashton, A. H. 1991. Experience and error frequency knowledge as potential determinants of audit expertise. *The Accounting Review* 66 (April): 218–39.

——, and R. H. Ashton. 1988. Sequential belief revision in auditing. *The Accounting Review* 63 (October) 623–41.

Ashton, R. H. 1974. An experimental study of internal control judgments. *Journal of Accounting Research* 12 (Spring): 143–57.

———. 1982. Discussion of "An assessment of laboratory experiments in accounting." *Journal of Accounting Research* 20 (Supplement): 102–07.

Bonner, S. E., and B. L. Lewis. (in press). Determinants of auditor expertise. *Journal of Accounting Research*.

Brehmer, B. 1990. Strategies in real-time, dynamic decision making. In *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, edited by R. M. Hogarth. Chicago: The University of Chicago Press.

Brunswik, E. 1956. *Perception and the Representative Design of Experiments*. Berkeley: University of California Press.

Buchman, T. A. 1985. An effect of hindsight on predicting bankruptcy with accounting information. *Accounting, Organizations and Society* 10, 267–85.

Cushing, B. E., and J. K. Loebbecke. 1986. *Comparison of Audit Methodologies of Large Accounting Firms*. Sarasota, FL: American Accounting Association.

Dawes, R. M., and B. Corrigan. 1974. Linear models in decision making. *Psychological Bulletin* 81 (No. 2). 95–106.

———, D. Faust, and P. E. Meehl. 1989. Clinical versus actuarial judgment. *Science* 243 (March): 1668–74.

de Groot, A. D. 1965. *Thought and Choice in Chess*. The Hague, The Netherlands: Mouton.

Einhorn, H. J., and R. M. Hogarth. 1975. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance* 13: 171–92.

———, and ———. 1978. Confidence in judgment: Persistence of the illusion of validity. *Psychological Review* 85 (No. 5): 395–416.

Frederick, D. M. 1991. Auditors' representation and retrieval of internal control knowledge. *The Accounting Review* 66 (April): 240–58.

Gibbins, M. 1982. Regression and other statistical implications for research on judgment using intercorrelated data sources. *Journal of Accounting Research* 20 (Spring): 121–38.

Goldman, A. 1978. Epistemics: The regulative theory of cognition. *Journal of Philosophy* 75 (No. 10): 509–523.

Hasher, L., and R. T. Zacks. 1979. Automatic and effortful processes in memory. *Journal of Experimental Psychology: General* 108 (No. 3): 356–88.

———, and ———. 1984. Automatic processing of fundamental information: The case of frequency occurrence. *American Psychologist* 39 (No. 12): 1372–88.

Herrnstein, R. J. 1990. Rational choice theory: Necessary but not sufficient. *American Psychologist* 45 (No. 3): 356–67.

Hogarth, R. M. 1981. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin* 90: 197–217.

———. 1986. Generalization in decision research: The role of formal models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-16*: 439–49.

———. 1987. *Judgement and Choice: The Psychology of Decision*. 2d ed. Chichester, England: Wiley.

———, and H. J. Einhorn. 1990. Order effects in belief updating: The belief-adjustment model. Unpublished manuscript, Graduate School of Business, University of Chicago.

———, B. J. Gibbs, C. R. M. McKenzie, and M. A. Marquis. (in press). Learning from feedback: Exactingness and incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Johnson, P. E., K. Jamal, and R. G. Berryman. 1989. Audit judgment research. *Accounting, Organizations and Society* 14: 83–99.

Joyce, E. J., and G. C. Biddle. 1981. Are auditors' judgments sufficiently regressive? *Journal of Accounting Research* 19 (Autumn): 323–49.

Kahneman, D., P. Slovic, and A. Tversky. 1982. Eds. *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Kleinmuntz, D. N. 1985. Cognitive heuristics and feedback in a dynamic decision environment. *Management Science* 31 (No. 6): 680–702.

———. 1990. Decomposition and the control of error in decision-analytic models. In *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, edited by R. M. Hogarth. Chicago: The University of Chicago Press.

Kiersey, G. F., and T. J. Mock. 1989. Verbal protocol research in auditing. *Accounting, Organizations and Society* 14: 133–51.

Krogstad, J. L., R. T. Ettenson, and J. Shanteau. 1984. Context and experience in auditors' materiality judgments. *Auditing: A Journal of Practice & Theory* 4 (Fall): 54–74.

Larkin, J. H., J. McDermott, D. P. Simon, and H. A. Simon. 1980. Expert and novice performance in solving physics problems. *Science* 208 (June): 1335–42.

Libby, R. 1985. Availability and the generation of hypotheses in analytical review. *Journal of Accounting Research* 23 (Autumn): 648–67.

——, and D. M. Frederick. 1990. Experience and the ability to explain audit findings. *Journal of Accounting Research* 28 (Autumn): 348–67.

McDaniel, L. 1990. The effects of time pressure and audit program structure on audit performance. *Journal of Accounting Research* 28 (Autumn): 267–85.

Lopes, L. L. 1987. Procedural debiasing. *Acta Psychologica* 64: 167–85.

McKenzie, C. R. M. 1990. A note on the accuracy of descriptive covariation strategies. Unpublished manuscript, University of Chicago, Department of Psychology.

Meehl, P. E. 1954. *Clinical versus Statistical Prediction: A Theoretical Analysis and Review of the Evidence*. Minneapolis: University of Minnesota Press.

Newell, A., and H. A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Oskamp, S. 1962. The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs*, Vol. 76, No. 28.

Peterson, C. R., and L. R. Beach. 1967. Man as an intuitive statistician. *Psychological Bulletin* 68: 29–46.

Saaty, T. L. 1980. *The Analytic Hierarchy Process: Planning Priority Setting, Resource Allocation*. New York: McGraw-Hill.

Schein, E. H. 1985. *Organizational Culture and Leadership: A Dynamic View*. San Francisco, CA: Jossey-Bass.

Shanteau, J. 1987. A psychologist looks at accounting research: Implications for doctoral programs. *A Passion for Excellence in Accounting Doctoral Education*. Sarasota, FL: American Accounting Association.

——. (in press). The psychology of experts: An alternative view. *Organizational Behavior and Human Decision Processes*.

Simon, H. A. 1957. *Models of Man*. New York: Wiley.

Slovic, P., and S. Lichtenstein. 1971. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance* 6: (No. 6): 649–744.

Spires, E. E. 1991. Auditors' evaluation of test-of-control strength. *The Accounting Review* 66 (Spring): 259–76.

Sterman, J. D. 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making environment. *Management Science* 35 (No. 3): 321–39.

Swieringa, R. J., and K. E. Weick. 1982. An assessment of laboratory experiments in accounting. *Journal of Accounting Research* 20 (Supplement): 56–101.

Thorngate, W. 1980. Efficient decision heuristics. *Behavioral Science* 25: 219–25.

Toda, M. 1962. The design of a fungus eater: A model of human behavior in an unsophisticated environment. *Behavioral Science* 7 (No. 2): 164–83.

Waller, W. S., and W. L. Felix, Jr. 1984. The auditor and learning from experience: Some conjectures. *Accounting, Organizations and Society* 9 (No. 3/4): 383–406.

——, and ——. 1987. Auditors' covariation judgments. *The Accounting Review* 62 (April): 275–92.

Weber, R. 1980. Some characteristics of free recall of computer controls by EDP auditors. *Journal of Accounting Research* 18 (Spring): 214–41.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 291–312

The Information Content of Annual Reports: A Price and Trading Response Analysis

William M. Cready
University of North Carolina at Chapel Hill
Patricia G. Mynatt
University of North Carolina at Charlotte

SYNOPSIS: This research examines and interprets the security market response around annual report release dates. Annual reports contain data beyond that reported in preliminary earnings announcements. While prior research has concentrated on the price response to annual report releases and the issue of information content with mixed results, this paper extends earlier work by employing both price and trading measures. The trading measures, which include number of transactions, size-stratified transactions, and mean transaction size, enable the analysis to address not only information content issues, but also social welfare issues in the sense of Lev and Ohlson (1982) and to coarsely identify what type of investor responds to the annual report by trading. Inferences as to the relative usefulness of annual reports to different investor groups are also made based on these trading measures.

A number of studies address, often indirectly, annual report informativeness by examining the price response accompanying the report's release. Of these, studies by Wilson (1987) and Lobo and Song (1989) suggest a price response to the earlier of the annual report or 10-K. However, other studies, including Foster et al. (1986), Mynatt (1988), and Bernard and Stober (1989), fail to detect such a price response. The informativeness of annual reports from a price-based perspective remains an open question and from a trading-based perspective it is an unaddressed question.

In this study, price response to the annual report is measured using both absolute value and squared unexpected returns. A market model ap-

We are grateful to Jens Stephans for his input in the initial stages of this paper. We also thank the referees for their helpful comments and insights.

Submitted November 1989.
Revisions received May and September 1990.
Accepted October 1990.

proach is used to estimate unexpected returns. A simulation analysis is included to substantiate the ability of these two metrics to detect a price response. Unexpected volume and unexpected number of transactions, both in general and for size-stratified trading, are estimated via market model regressions. Unexpected mean transaction sizes are estimated from a temporal drift regression.

No evidence of a price response and little evidence of a volume of shares response at annual report dates is found. Number of transactions, however, increases significantly around annual reports, peaking four to five trading days after the annual report release date. This trading response suggests that investors find annual reports informative. The simultaneous absence of an economically significant price response is interpreted as indicating that annual reports provide social welfare value in the sense proposed by Lev and Ohlson (1982).

Most notably, the analysis of trading response stratified by transaction size suggests that the trading response occurs mostly within the smallest size strata (i.e., 100 and 200 share transactions, transactions of less than $30,000 in value). Consistent with this, mean transaction sizes during the annual report period are less than expected. This is contrary to mean transaction size behavior at earnings announcement dates (Cready 1988) and suggests that annual reports are of greater value to less wealthy individual investors relative to wealtheir individual and institutional investors. Such a finding is consistent with Hakansson (1977) in that it suggests that "small" investors rely on the public information system (i.e., the annual report) while "large" investors rely more on predisclosure information in making investment decisions.

Key Words: *Annual report, Trading response, Information content, Small versus large investors.*

Data Availability: *The sample of firms together with their annual report release dates is available from the authors. The remaining data can be gathered from sources identified in the article.*

THE remainder of this paper is organized as follows. The next section discusses trading activity as a measure of information content, as a measure of social welfare, and as a measure of relative usefulness among different investor groups. The third section develops the research design, including the research hypotheses and sample selection. The fourth section presents the empirical findings and the fifth section presents conclusions.

I. Trading Activity as a Market Response Measure

In addition to trading volume, this study employs total and size-stratified numbers of transactions in measuring trading activity. Furthermore, average transaction size (volume divided by transactions) is employed as a measure of relative trading activity across investor groups.

Beaver (1968) introduced volume as a means of measuring the degree of consensus among investors with respect to a given information disclosure. He found that trading volume increases significantly at annual preliminary announcement dates, implying that there may be a lack of consensus across market participants in interpreting earnings data. Verrecchia (1981) questions such an interpretation of volume results and demonstrates that information can induce trading even in the presence of consensus interpretation. Hakansson et al. (1984) provide a precise set of necessary conditions for a no-trading response to occur with respect to an information release. Since both Verrecchia and Hakansson et al. suggest it is improbable that an information release will satisfy all of these necessary conditions, trading activity appears better interpreted not as degree of consensus but rather as evidence of information content.

Trading and Social Welfare

Lev and Ohlson (1982) propose using trading as a social welfare indicator. They show that trading activity induced by an information release evidences a strict social welfare benefit, provided that the market exhibits price invariance with respect to the information release. Price invariance, meaning that "security prices are the same whether or not the informational item (or output of the information structure) is known by all individuals in the economy" (Lev and Ohlson 1982, 295–96), is a crucial assumption. The welfare interpretation arises under price invariance because the information item's release does not cause price changes and therefore cannot make anyone worse off (e.g., by lowering endowments or raising acquisition costs), while the presence of trading activity suggests that some investors are trading to make themselves better off.

Price invariance, though, constitutes a major stumbling block to empirical applications of the Lev and Ohlson framework. Most information releases affect price as well as trading; indeed, one finds researchers assuming, often implicitly, that a release lacks information content if no price change is observed. From a statistical perspective, price invariance is difficult to prove. Establishing price invariance involves failing to reject the null hypothesis of no information-induced change in price, a far more problematic task than rejecting this same hypothesis. Indeed, absent some tolerance for the existence of some small, but subjectively judged insignificant price change, one cannot make meaningful probabilistic inferences as to the absence of a price change to a given information release.[1]

Trading Measures

Prior trading-based studies almost uniformly employ volume of shares in measuring trading activity (e.g., Beaver 1968; Morse 1981; Pincus 1983; Bamber 1986, 1987). However, recent studies by Howard (1982) and Cready (1988) also use number of transactions to measure trading. This study reports both measures since the frameworks of Hakansson et al. (1984) and Lev and Ohlson (1982) are robust with respect to measurement of trading. We, however, prefer the number of transactions measure for several reasons.[2] First, since number of transactions counts each trade, and thus each trader, equally, it may be more appropriate as a social welfare measure than trading volume, where each transaction is implicitly weighted by its size (i.e., large block transactions have a relatively large impact while small transactions have a relatively small impact). Second, in a simulation study Cready and Ramanan (1990a) find that statistical tests employing number of transactions appear to be considerably more powerful than tests employing trading volume in measuring unexpected trading activity.[3] Third, transactions can be stratified by size, as in Cready (1988), providing insights into characteristics of those investors responding to annual reports.

---

[1] Tolerance for some non-zero price change is necessary because assessing the likelihood that the null hypothesis is false requires a precise statement of the (true) alternative hypotheses. Specifically, in this paper it is necessary to know a precise value by which the actual price response might differ from zero in order to generate likelihoods that the null is false (power curves) conditional on achieved test statistic values.

Size-Stratified Trading and Mean Transaction Size

In evaluating welfare/equity issues, Lev (1988) emphasizes the importance of focusing on the individual investor and investor classes. Cready (1988) proposes the use of transaction size as a measure of investor-type with very small transactions (100 and 200 shares) identified as small individual investor trading, relatively larger transactions (300 to 900 shares) identified as large (wealthier) individual investor trading, and large transactions (2,000 or more shares) identified as institutional investor trading. He proposes the use of mean transaction size as a measure of relative trading activity across these groups.

Size-stratified and mean transaction size analyses of annual report trading possess special interest in the context of the theoretical work of Hakansson (1977). Hakansson develops a model in which firms make user-costly interim disclosures prior to the release to all investors of "annual" reports. When investor groups differ in information acquisition abilities and/or resources (e.g., wealth), distinct patterns of information acquisition emerge. Investors with high information acquisition skill or large resources place greater emphasis on interim disclosure information relative to low information acquisition skill and low resource investors. Consequently, relatively less emphasis is placed on the ultimate "annual" report disclosure. An examination of size-stratified trading and mean transaction size behavior offers insight into the descriptive validity of the Hakansson model.

II. Research Design

The research design consists of three stages. The first stage examines the price response to the annual report using unexpected returns at annual report release dates. The second stage examines the trading response to the annual report using trading volume, number of transactions, and size-stratified trading. The third stage, which is conditional on the presence of a trading response at annual report dates (i.e., stage two), examines relative trading at report dates by investor groups.

---

[2] Transactions measures are not without drawbacks. Some theories of trading, particularly those based on informational asymmetries (see Jang and Ro 1989 for a review), map much more closely to volume than to transactions in terms of the relevant trading measure. Transactions data are also not as widely available as volume data although this state of affairs is changing.

[3] For instance, Cready and Ramanan (1990a) find that across 500 portfolios of 50 firms each, an induced 20 percent one-day increase in volume is detected at the five percent level 50 percent of the time, while an induced 20 percent one-day increase in number of transactions is detected at the five percent level over 90 percent of the time.

Unexpected Returns Design

Price response to the annual report is measured using standardized absolute and squared unexpected returns. Both procedures remove directionality from the returns and allow aggregation of the measure across firms with positive and negative unexpected returns.

The market model is estimated for each firm using daily data and is used to form daily unexpected return (UR).[4] Squared standardized unexpected return (SUR) is then estimated for each firm on a daily basis following Patell (1976). The expected value of SUR is equal to 1.0 in the case of no price response.

Rohrbach and Chandra (1989) find that tests based on the absolute value of returns may be more powerful than tests based on squared returns when returns are not normally distributed. Other studies, such as Marais (1985), find that returns indeed are not normally distributed. Accordingly, we also employ absolute value of unexpected return ($ABUR$) in evaluating price invariance. $ABUR$ is estimated on a daily basis as:

$$ABUR_{it} = \left(\frac{|UR_{it}| - \overline{ABUR}_i}{SABR_i} \right) \tag{1}$$

where:

$\overline{ABUR}_i$: the mean absolute value of the residuals for firm i from the market model estimation period; and

$SABR_i$: the standard deviation of the absolute values of the residuals for firm i from the market model estimation period.

Trading Response Design

Trading response is measured using both volume of shares and number of transactions. Furthermore, the trading response within transaction size classes is estimated by examining size-stratified trading activity.

Expected trading volume, $\hat{PV}$, is estimated from the following time-series regression:

$$PV_{it} = a_i + b_i \ln(VMKT_t / OSMKT_t) + e_{it}, \tag{2}$$

where:

PV_{it}: the natural logarithm of daily trading volume for firm i in shares (V) divided by number of outstanding shares for firm i, both on day t, times 100 plus .000255;[5]

---

[4] Ordinary Least Squares (OLS) is used to estimate the parameters of the model. Brown and Warner (1985) and Jain (1986) support the use of OLS for estimating unexpected returns even in the presence of nonsynchronous trading. Nonsynchronous trading is not expected to be significant for this sample given the sample selection criterion that stipulates a minimum average daily number of trades of eight during the estimation period. The model is estimated over a 200-day period, day −210 to day −11, relative to the prior preliminary earnings announcement. The model is estimated using the CRSP equally weighted market index.

[5] Outstanding shares data are obtained from the daily CRSP master which incorporates all changes in number of outstanding shares, including stock dividends and splits. Volume is traditionally standardized by outstanding shares (e.g., Beaver 1968; Bamber 1986, 1987). Cready and Ramanan (1990b), Ajinkya and Jain

$VMKT_t$: total daily trading volume for all NYSE firms on day t times 100;
$OSMKT_t$: total outstanding shares for all NYSE firms on day t;
a_i, b_i: the estimated intercept and slope terms; and
e_{it}: the error term.

The regression is estimated subject to a first-order serial correlation (see Pincus 1983; Richardson et al. 1986).[6] Unexpected volume (UV) is then estimated on a daily basis as:

$$UV_{it}=(PV_{it}-\widehat{PV}_{it})/\widehat{STV}_i, \tag{3}$$

where $\widehat{STV}_i$ is the standard deviation of the residuals for firm i from equation (2) and $\widehat{PV}$ is obtained from the structural portion of equation (2).[7]

Number of transactions, like volume, is a measure of trading, and if marketwide factors influence volume then they should also influence number of transactions. Thus, like Cready (1988) and Cready and Ramanan (1990a), a market model approach is used in estimating expected number of transactions.[8] The specific regression model employed is:

$$TR_{it}=a_i+b_i\ell n(TRMKT_t)+e_{it}, \tag{4}$$

where:

TR_{it}: the natural logarithm of the daily number of transactions occurring in firm i's stock on day t, plus 1;[9]
$TRMKT_t$: the number of transactions occurring in all NYSE stocks on day t;
a_i, b_i: the estimated intercept and slope terms; and
e_i: the error term.

Equation (4) is also estimated subject to a first-order serial correlation. Unexpected number of transactions is then estimated as:

$$UTR_{it}=(TR_{it}-\widehat{TR}_{it})/\widehat{STR}_{it}, \tag{5}$$

where $\widehat{STR}_i$ is the standard deviation of the residuals from equation (4), and $\widehat{TR}$ is obtained from the structural portion of equation (4).[10]

(1989), Richardson et al. (1986), and Pincus (1983) suggest the use of the logarithmic transformation. The constant, 0.000255, added in order to preclude taking the logarithm of zero, is arrived at by dividing the weekly constant of 0.001275 used by Richardson et al. (1986) by five.

[6] The autoregression estimation employs Prais-Winsten estimators in an approach very similar to the Cochrane-Orcutt procedure.

[7] Only the structural portion is used because of the multiday nature of the anticipated trading response. Specifically, if positive unexpected trading occurs on event day t and the serial correlation is positive (as is the case here), expected trading from the full model on day $t+1$ is biased upward. Of course, failure to account for the serial correlation raises the noise level of the unexpected trading measure.

[8] The analysis of Cready and Ramanan (1990a) suggests that a market model approach measurably outperforms a mean approach (i.e., expected number of transactions equals their nonevent period mean) in terms of rejection percentages achieved. They also find that employing a market model approach eliminates the problems with overstated rejection frequencies associated with temporal clustering of event dates (i.e., it is robust to cross-sectional dependence).

[9] Cready and Ramanan (1990a) suggest the appropriateness of logarithmic transformations. The occurrence of zero transactions in some stocks on some days dictates the addition of a constant. The choice of one, while consistent with Cready and Ramanan, is *ad hoc.*

[10] Cready and Ramanan (1990a) find that UTR comes very close to possessing a normal distribution.

The regression equations for size-stratified trading follow the same form as equation (4), differing in that the number of transactions counted in the dependent variable is limited to those that meet the share size requirements of each stratum. Size strata employed are transactions of less than \$30,000 (*TV*030), transactions of \$100,000 or more (*TV*100), transactions of 100 and 200 shares (*TR*12), transactions of 300 to 900 shares (*TR*39), transactions of 100 to 900 shares (*TR*19), and transactions of 2,000 or more shares (*TR*2000). *UTR*12 is interpreted as a measure of unexpected small (low wealth) individual investor trading, *UTR*39 as a measure of unexpected large (high wealth) individual investor trading, *UTV*030 and *UTR*19 as measures of unexpected individual investor trading, and *UTV*100 and *UTR*2000 as measures of unexpected institutional investor trading.[11]

Relative Trading Response Design

Relative trading activity is measured using mean transaction size in shares. Following Cready (1988), two such mean transaction-size measures are used, one for transactions of 900 or fewer shares (*MTR*9) and one for all transactions (*MTR*). *MTR*9 measures relative trading activity by wealthy versus less wealthy individual investors while *MTR* is sensitive to the wealth of all investors, including institutions. Since institutions engage in predominantly large transactions and individuals engage in predominantly small transactions, MTR should be viewed as a measure of institutional relative to individual investor trading. One drawback to *MTR* is its sensitivity to extremely large-number-of-share transactions. A single large block trade can, in and of itself, lead to a very large *MTR* observation. The severity of this problem is in inverse proportion to the number of transactions over which MTR is calculated. It is partly for this reason that an average number of transactions per day restriction is placed on the sample. Expected values for *MTR* (i.e., $\widehat{MTR}$) and MTR9 (i.e., $\widehat{MTR}9$), are estimated in the manner introduced by Cready (1988), using regressions of the form:

$$MTR_{it} = a_i + b_i(DAY_t) + e_{it} \tag{6}$$

and

$$MTR9_{it} = c_i + d_i(DAY_t) + e_{it} \tag{7}$$

where:

MTR_{it}: the mean transaction size in shares for transactions occurring in firm *i*'s stock on day *t*;
$MTR9_{it}$: the mean transaction size in shares for transactions of 900 or fewer shares occurring in firm *i*'s stock on day *t*;
DAY_t: the trading day per the daily CRSP tape;
a_i, c_i: regression-specific estimated intercept term for firm *i*;
b_i, d_i: regression-specific estimated slope term for firm *i*; and
e_{it}: the regression-specific error term.

---

[11] The transaction share size strata correspond to those employed in Cready (1988). The two transaction value strata are *ad hoc*. As cited in Cready (1988), *Shareownership 1983* provides empirical evidence for a positive relation between transaction size and investor wealth.

Table 1
Unexpected Trading and Relative Trading Measures

Unexpected Volume	
UV	All transactions
Unexpected Number of Transactions	
UTR	All transactions
UTV030	Transactions of less than $30,000
UTV100	Transactions of $100,000 or more
UTR12	Transactions of 100 and 200 shares
UTR39	Transactions of 300 to 900 shares
UTR19	Transactions of 100 to 900 shares
UTR2000	Transactions of 2,000 or more shares
Relative Trading—Unexpected Mean Transaction Size	
UMTR	All transactions
UMTR9	Transactions of 900 or fewer shares

Unexpected mean transaction sizes are then estimated as:

$$UMTR_{it}=MTR_{it}-\widehat{MTR}_{it}, \qquad (8)$$

and

$$UMTR9_{it}=MTR9_{it}-\widehat{MTR}9_{it}. \qquad (9)$$

All trading- and relative-trading-response measures are summarized in table 1. All trading models are estimated using 201 days of data centered on the annual report release date. However, trading days where the number of outstanding shares for a sample firm differs by more than five percent from the number of outstanding shares at the annual report release date are excluded to enhance intertemporal comparability of trading activity.

Data

The returns and outstanding share data used in this study are obtained from CRSP while the volume and transaction data are obtained from daily transaction records for all NYSE-listed securities provided by Francis Emory Fitch, Inc. The Fitch records list the date, time, price, and number of shares for all round lot trades.[12] The Fitch data employed in this study cover four years (1/1/81 through 12/31/84). Consistent with Cready (1988), regional exchange trading is excluded because of apparent irregularities and its relatively small contribution to overall trading. Daily size stratified trading and mean transaction sizes are calculated excluding the first fifteen minutes of trading in order to mitigate the effects of overnight order accumulations.

---

[12] Odd lot data are not provided by Fitch and thus are necessarily excluded from the analysis. It is not clear, however, whether the inclusion of odd lot data would be desirable even if it were available.

Table 2
Hypotheses
(In Null Form)

Price	*Aggregate Trading Response*	*Relative Trading Response*
H1.1: $\overline{SUR} \le 1$.	H2.1: $\overline{UV} \le 0$.	H3.1: $\overline{UMTR} = 0$.
H1.2: $\overline{ABUR} \le 0$.	H2.2: $\overline{UTR} \le 0$.	H3.2: $\overline{UMTR9} = 0$.
	H2.3: $\overline{UTV030} \le 0$.	
	H2.4: $\overline{UTV100} \le 0$.	

Hypotheses

The research design encompasses three major hypotheses. The hypotheses are presented in the null form in table 2. They are tested using cross-sectional *t*-tests.[13]

The first hypothesis (H1) constitutes a test of price invariance. Failure to reject either hypotheses H1.1 or H1.2 is consistent with price invariance and the results of Foster et al. (1986), Mynatt (1988), and Bernard and Stober (1989). A failure to reject does not by itself allow us to make probabilistic conclusions regarding price invariance. We do make such conclusions, but they necessitate allowing for the possible occurrence of "small" price changes (e.g., less than one percent, on average).

The second hypothesis (H2) concerns trading activity at annual report dates. Unexpectedly high trading activity at these dates is evidence of: (1) information content (Verrecchia 1981; Hakansson et al. 1984); and (2) social welfare value, under the assumption of price invariance (Lev and Ohlson 1982). We conduct four empirical tests of the presence of such unexpected trading activity. Hypotheses H2.1 and H2.2 refer to aggregate trading. Rejection of one or both of these hypotheses suggests increased trading, consistent with annual reports possessing information content and social welfare value (assuming hypothesis H1 is not rejected).

Since report-induced trading may be limited to only certain types of investors, we also test for the presence of unexpected trading activity among individual and institutional investors using value-stratified trading.[14] Specifically, with respect to individual trading, we test the hypothesis H2.3. With respect to institutional trading, we test the hypothesis H2.4.

The third hypothesis (H3) concerns relative trading activity by investor groups at annual report dates. Hakansson (1977) proposes that annual report acquisition/utilization varies inversely with investor resources suggesting that small investor transactions (i.e., small share size trades) should predominate in response to annual reports. However, if annual report information value increases with an investor's resources as

---

[13] Cready and Ramanan (1990a) find that a cross-sectional test approach performs as well as alternative time-series-based designs for both transactions and volume.

[14] We do not consider share-size-stratified trading (e.g., *TR*19, *TR*12) at this stage of the analysis because: (1) the share-size-stratified results are for a smaller price-restricted sample, and (2) there is little difference, in terms of making an inference as to the simple presence of unexpected trading, between the share-size-stratified and the value-stratified results.

Cready (1988) concludes is the case with respect to earnings announcements, then large investor transactions should predominate. Given this ambiguity in the expected direction of the relative trading response, we employ nondirectional hypotheses in examining relative trading. We examine relative institutional versus individual trading at report dates by testing hypothesis H3.1. Rejection of this hypothesis suggests differential relative trading levels for individual versus institutional investors at annual report dates. If $\overline{UMTR}$ is negative (positive), it suggests that individual trading is relatively heavier (lighter) than institutional trading.

We examine the relation between individual investor wealth and trading at report dates by testing hypothesis H3.2. Rejection of this hypothesis would suggest a relation between relative trading and wealth among individual investors at annual report dates. If $\overline{UMTR9}$ is negative (positive), rejection suggests that the trading response among individual investors decreases (increases) with wealth.

Following Cready (1988), hypotheses H3.1 and H3.2 are tested on a reduced sample of firms. Specifically, in order to maintain a degree of cross-sectional uniformity in share price (a major determinant of mean transaction size) only those sample firms with mean closing prices (over the 201 days centered on the report date) between $10 and $40 are included in this reduced sample.

Sample

All NYSE-listed firms with fiscal year-ends falling between 6/1/81 and 5/31/84 (i.e., fiscal years 1981 through 1983) constituted the set of initial candidates for inclusion in the sample. We randomly selected 320 firms from each of the three fiscal years. Microfiche copies of annual reports (produced by Disclosure, Inc.) were located in library files for 822 of these firms over the three years.[15] We then deleted reports if: (1) the firm's returns were missing or the firm could not be located on CRSP; (2) the subsequent first quarter's earnings announcement date is unavailable; (3) the annual report date is within ten trading days of the first quarter announcement date;[16] (4) the annual report date is within 15 trading days of the prior annual preliminary earnings announcement date; (5) the number of outstanding shares on fewer than 121 trading days in the combined estimation and announcement period (i.e., 100 nonannouncement period observations) differ by less than five percent from the number of shares outstanding on the annual report release date; and (6) trading activity averages less than eight transactions per day over the estimation period.

The first sample restriction is dictated by the use of returns data in examining the price response. Restrictions two through four minimize the potential confounding influence of annual and first-quarter preliminary earnings announcement trading on our measures of annual report trading. The fifth restriction guarantees that the time-series regression is based on a minimum of 100 comparable nonannouncement period observations, while the sixth restriction eliminates infrequently traded firms from the sample. Application of these criteria resulted in a sample of 563 annual reports dates as detailed in table 3.

---

[15] This is the same initial sample used by Mynatt (1988).

[16] The annual report date is defined as the earlier of the SEC or Disclosure, Inc., stamped receipt date, as available on the microfiche copies of the annual reports provided by Disclosure, Inc. The announcement period is defined as trading days −9 to +10 relative to the annual report receipt date.

Table 3
Annual Report Date Sample Selection Procedure (Fiscal Years 1981 through 1983)

Initial Sample of 320 Firms Per Year Obtained from COMPUSTAT	960
Unable to Locate Annual Report Date	(138)
Missing or Unable to Locate on CRSP Tape	(81)
Missing Subsequent First Quarter Announcement Date	(3)
Report Date Within 10 Days of Subsequent First Quarter Announcement Date	(52)
Report Date Within 15 Days of Annual Preliminary Earnings Announcement Date	(31)
Fewer than 121 Trading Observations Satisfy Outstanding Shares Criterion[a]	(30)
Fewer than 8 Transactions Per Day	(62)
Final Sample	563

[a] Trading days within 100 days of the annual report date for a given firm are excluded if the number of outstanding shares on that date per CRSP differs by more than five percent from the number of outstanding shares on the annual report date for that firm.

Also, as noted earlier, when trading measures are based on transaction share size (e.g., *TR*19, *MTR*) rather than transaction value, the only firms included are those with average closing prices between $10 and $40 over the 201 days centered on the report date. This restriction results in the loss of an additional 141 firm report dates when employing these measures.[17]

The unrestricted sample consists of 177 reports for fiscal year 1981, 200 reports for fiscal year 1982, and 186 reports for fiscal year 1983. Approximately 30 percent of the reports are for non-12/31 fiscal year firms, so event-date clustering is not a significant problem. Table 4 provides descriptive statistics on returns, trading activity, and firm size for the sample.[18] Table 4 also provides information on annual report timing.[19]

[17] There are no substantive differences between any of the reported results when based on just this 422 report subsample.

[18] In table 3, average transaction sizes in shares are somewhat higher than those reported in Cready (1988), reflecting the lack of a mean price exclusion for this sample and its more recent dating (Cready finds that mean transaction size drifts upward over time). Average volume of 0.1985 percent is also somewhat higher than that reported in prior studies (e.g., in Bamber 1986 it equals 0.1064 percent). This reflects the fact that these are largely post-August 1982. Post-August 1982 volume is considerably greater than pre-August 1982 volume due to the start of the 13 August 1982 bull market. Consistent with other studies, a sizable positive serial correlation exists for daily percentage volume and number of transactions. Mean R^2 for the percentage volume first-order autoregression equals 0.163. While this is low relative to number of transactions, it is high compared to other studies where this statistic is reported (e.g., 11 percent in Ajinkya and Jain 1989). This again may reflect the time period covered in this study. Mean R^2 for number of transactions first-order autoregression, at 0.352, is quite high relative to volume. This is consistent with the results of Cready and Ramanan (1990a), who find number of transactions R^2s to be much higher than percentage volume R^2s.

[19] While annual reports on average tend to precede 10-Ks, this is not the case for nearly 40 percent of our sample. However, only ten percent of the annual report dates are more than five trading days after the 10-K date. The overlapping of dates for the 10-K and annual report raises the issue of which event (the 10-K or annual report release) is the underlying cause of any observed market response. Since the focus of this paper is trading we believe that the more sensible interpretation of an observed trading response is as a reaction to the widely disseminated annual report, not to the selectively disseminated 10-K. Furthermore, in an analysis not directly reported here, we restrict the sample to only those annual reports that precede the 10-K by five or more trading days and that obtain similar results.

Table 4
Sample Description[a]

Variable	*Mean*	*Standard Deviation*	*Minimum*	*Maximum*
Average Daily Trading Volume in Shares	67,958	87,635	30,333	723,006
Mean Transaction Size[c] in Shares	1,209	581	278	3,140
Mean Transaction Size,[c] Transactions ≤ 900 Shares	276.2	25.1	208.6	356.3
Average Daily Percentage Volume	0.1985%	0.1415%	0.0412%	1.2533%
Average Daily Number of Transactions	45.87	51.33	8.02	556.82
Average Number of Transactions <$30,000	31.72	32.84	4.58	393.84
Average Number of Transactions >$100,000	4.31	10.42	0.0	167.76
Average Daily Return	0.00051	0.00164	−0.0077	0.00687
Market Value at Year-End	$1,506 MM	$3,549 MM	$246 MM	$57,982 MM
Number of Days Following Preliminary Announcement	30.1	7.8	15	51
Number of Days Before First Quarter Announcement	23.5	7.8	11	58
Annual Report Trading Day Less 10-K Trading Day[b]	3.4	8.4	−31	27
Volume Model:[d]				
R^2	0.163	0.105	0.001	0.624
Serial Correlation coefficient	0.252	0.129	0.687	−0.087
Transactions Model:[e]				
R^2	0.352	0.156	0.046	0.850
Serial Correlation coefficient	0.436	0.159	0.000	0.835

[a] The sample consists of 563 firm report dates, 177 of which are 1981 fiscal years, 200 of which are 1982 fiscal years, and 186 of which are 1983 fiscal years.

[b] Annual Report and 10-K dates are each the earlier of their SEC or Disclosure, Inc., stamped receipt dates. Data Available for 559 firms only, Annual Report Date Precedes 10-K for 350 of these.

[c] Based on the 422 Firms in the Sample With Mean Closing Prices Between $10 and $40.

[d] This is a first-order autoregressive model, estimated over days −100 to +100 relative to the report date, where the logarithm of the percentage of a firm's outstanding shares is the dependent variable and the logarithm of the percentage of outstanding shares traded for all NYSE firms is the independent variable.

[e] This is a first-order autoregressive model, estimated over days −100 to +100 relative to the report date, where the logarithm of the number of transactions in a firm's stock is the dependent variable and the logarithm of the total number of transactions for all NYSE firms is the independent variable.

III. Empirical Findings

Price Response

Table 5 reports both daily mean squared and absolute value standardized unexpected returns for days −9 to +10 relative to the annual report date and for two multiday periods. The squared returns price response measure is insignificantly different

Table 5

Mean Squared Standardized Unexpected Returns and Mean Absolute Value Standardized Unexpected Returns[a] at Annual Report Dates

	Squared Unexpected Returns				*Absolute Value Unexpected Returns*			
Day(s) Relative to Report Date[b]	*Mean*	*Standard Deviation*	t	P	*Mean*	*Standard Deviation*	t	P
−9	1.13	3.21	0.95	0.17	0.035	1.13	0.73	0.23
−8	1.07	2.84	0.60	0.27	0.016	1.09	0.35	0.37
−7	1.22	4.89	1.06	0.14	0.040	1.22	0.78	0.22
−6	0.98	2.05	−0.19	0.58	−0.022	1.03	−0.50	0.69
−5	1.19	4.28	1.05	0.15	0.001	1.22	0.02	0.49
−4	0.99	1.92	−0.02	0.51	0.018	1.01	0.42	0.34
−3	1.16	3.95	0.94	0.18	0.045	1.13	0.94	0.18
−2[c]	1.31	7.47	0.99	0.16	0.002	1.28	0.04	0.48
−1	1.00	1.95	0.03	0.49	−0.024	1.05	−0.54	0.70
0	1.05	2.41	0.52	0.30	0.012	1.07	0.27	0.40
+1	1.00	1.96	0.01	0.49	0.001	1.03	0.02	0.49
+2	0.97	1.76	−0.45	0.67	−0.010	1.01	−0.23	0.59
+3	1.14	3.01	1.12	0.13	0.021	1.15	0.44	0.33
+4	1.06	2.41	0.55	0.29	0.006	1.07	0.13	0.45
+5	1.05	2.07	0.59	0.28	0.005	1.07	0.12	0.45
+6	1.05	3.60	0.34	0.37	−0.069	1.12	−10.46	0.93
+7	0.99	2.22	−0.15	0.56	−0.038	1.05	−0.86	0.81
+8	0.94	2.02	−0.70	0.76	−0.086	1.04	−1.95	0.97
+9	0.92	1.66	−1.08	0.86	−0.032	0.97	−0.79	0.78
+10	1.09	2.49	0.87	0.19	0.002	1.11	0.05	0.48

	Cumulative Squared Unexpected Returns				*Cumulative Absolute Value Unexpected Returns*			
	Daily Mean	*Standard Deviation*	t	P	*Daily Mean*	*Standard Deviation*	t	P
−1–+1	1.02	1.45	0.31	0.38	−0.003	0.72	−0.12	0.55
−2–+2	1.07	1.85	0.86	0.20	−0.004	0.60	−0.14	0.56

[a] Squared standardized unexpected returns are computed to have an expected value of 1.0 in the case of no price response. Absolute value standardized unexpected returns are computed to have an expected value of zero in the case of no price response.

[b] The report date is the earlier of the SEC or Disclosure, Inc. stamp date.

[c] One day −2 standardized return is abnormally large, which accounts for the increased standard deviations on day −2. Excluding this return greatly reduces both the standard deviations and means for day −2 as well as the day −2 to +2 period with no appreciable effect on the reported *t*'s.

from 1.0 for each day and both multiday periods examined. The absolute value price response measure is insignificantly different from zero for each day and both multiday periods examined;[20] hypothesis H1 cannot be rejected.[21] These results are consistent with Bernard and Stober (1989), Foster et al. (1986), and Mynatt (1988) and provide support for price invariance with respect to annual reports.

---

[20] The returns analysis was replicated using the value-weighted rather than the equally-weighted CRSP market returns index with similar results not reported here.

[21] Alternatively, the day 0–+1 $\overline{SUR}$ and $\overline{ABUR}$, where days 0 and +1 are relative to the preliminary earnings announcement, are 1.41 and 0.314, respectively. The first is significantly different from 1.0 at the 0.0001 level, and the second is significantly different from 0 at the 0.0001 level. These results are consistent with previous evidence of price movement associated with the preliminary earnings announcement.

In a separate simulation analysis conducted over 250 portfolios of 563 observations each (described in greater detail in appendix A), we evaluated the significance of the table 5 multiday P-values in terms of the likelihood that larger P-values would occur in the presence of various one-day induced return effects.[22] When employing a five-day window in detecting an induced one-day return effect of one percent, an $\overline{SUR}$ P-value in excess of 0.3 (table 5 reports a P-value of 0.45) occurred less than ten percent of the time, while for the more powerful *ABUR* measure a P-value in excess of 0.5 (table 5 reports a P-value of 0.56) never occurred. Indeed, for $\overline{ABUR}$ a P-value in excess of 0.5 occurs less than five percent of the time in the presence of a 0.75 percent induced return effect, and less than 25 percent of the time in the presence of a 0.5 percent induced return effect. Employing a three-day window lowers these likelihoods further. For a three-day window an induced one-day return effect of 0.75 percent results in an $\overline{SUR}$ P-value in excess of 0.3 (table 5 reports a P-value of 0.38) less than ten percent of the time and an $\overline{ABUR}$ P-value in excess of 0.5 (table 5 reports a P-value of 0.55) less than one percent of the time.

Unexpected Trading Activity

Table 6 reports daily and multiday mean unexpected percentage trading volume (*UV*) and mean unexpected numbers of transactions (*UTR*) for days −9 through +10 relative to annual report dates.[23] The multiday periods examined are wider than those employed in examining returns because report-induced trading is not expected to occur in its entirety immediately upon the release of the annual report. Efficiency of the market with respect to an information release refers to aggregate market behavior (price) and not to individual traders. Table 7 reports mean unexpected numbers of transactions among institutional (*UTV*100) and individual (*UTV*030) investors. The *UV* measure (table 6) is marginally significant (ten percent level) only on day 0; the *UTR* measure is significant at the ten percent level on days −2, +1, and +10; at the five percent level on days 0, +3, +6, and +7; and at the one percent level on days +4 and +5. The *UTR* measure is also significant at the five percent level or better in all three multiday periods examined. Thus, while we cannot reject hypothesis H2.1 we do reject hypothesis H2.2 and conclude that a trading response accompanies the annual report.

Table 7 reports the *UTV*100 measure to be significant only on day −1, and barely at the ten percent level, meaning we cannot confidently reject hypothesis H2.4. However,

---

[22] The no inducement (zero percent inducement) rejection percentages reported in appendix A also indicate that cross-sectional tests employing $\overline{SUR}$ and $\overline{ABUR}$ are prone to overreject the null hypothesis. Overrejections occur with respect to $\overline{SUR}$ because, as noted in Marais (1985), its empirical expected value slightly exceeds one (it averages around 1.05 in the simulation). It occurs with respect to $\overline{ABUR}$ because $\overline{ABUR}$ is variance-sensitive and is based on the difference between the absolute value of prediction error and the mean absolute value of estimation period residuals. Since prediction error variances slightly exceed residual variances, $\overline{ABUR}$ should slightly exceed zero (on average it exceeds zero by about 0.003 in the simulation).

[23] Only the structural portion of the autoregressive model is used for these results because of the anticipated multiday nature of the trading response. In results, not reported here, where the full autoregressive model including the estimated first-order autocorrelation term is used, the results, while somewhat different from those reported in table 4, still lead to rejection of hypothesis H2.2. In further results, not reported here, event day dummy variables were employed in estimating the original autoregressive models. Cross-sectional means of standardized dummy variable coefficients lead to a set of results virtually identical to those reported in table 4.

Table 6
Mean Unexpected Volume and Number of Transactions at Annual Report Dates

Day(s) Relative to Report Date[a]	*Number of Reports*[b]	*Mean Actual Less Expected Volume*		*Mean Actual Less Expected Number of Transactions*	
		Unstand.[c]	*Stand.*[d]	*Unstand.*[e]	*Stand.*[f]
−9	535	0.0003	0.0052	0.663	−0.014
−8	541	0.0102	0.0510	0.243	0.046
−7	547	0.0128	0.0503	−0.638	−0.009
−6	555	0.0113	−0.0007	−0.702	0.027
−5	562	0.0016	−0.0236	−0.902	−0.023
−4	563	−0.0030	0.0010	−0.813	0.009
−3	563	0.0021	−0.0145	−1.039	−0.014
−2	562	−0.0048	0.0004	0.397	0.064†
−1	563	−0.0051	−0.0171	1.104	0.026
0	563	0.0077	0.0577†	0.931	0.107‡
+1	563	−0.0032	−0.0138	0.214	0.076†
+2	563	0.0029	−0.0135	−0.402	0.059
+3	563	−0.0019	0.0072	0.983	0.102‡
+4	562	0.0044	0.0165	1.596	0.123*
+5	562	0.0064	0.0476	1.330	0.157*
+6	562	0.0079	0.0059	1.178	0.099‡
+7	547	0.0063	0.0249	0.672	0.105‡
+8	527	−0.0086	−0.0783	−0.612	0.018
+9	519	−0.0031	−0.0084	−0.385	0.022
+10	505	0.0059	0.0118	0.150	0.074†
−2–+2	563	−0.003	0.014	2.24	0.333‡
−4–+5	563	0.006	0.071	4.29	0.709*
−9–+10	478	0.099	0.222	5.38	1.23*

† Significant at the ten percent level (one-tailed cross-sectional *t*-test).

‡ Significant at the five percent level (one-tailed cross-sectional *t*-test).

* Significant at the one percent level (one-tailed cross-sectional *t*-test).

[a] The annual report date is the earlier of the SEC or Disclosure, Inc., stamp date.

[b] Days within ten days of the preliminary announcement date or five days of the first quarter announcement date are deleted, resulting in the reduced sample sizes on days −9 through −6 and +7 through +10. The loss of one firm on day −2 is because of missing trading data for that firm on day −2. The loss of one firm for days +10 to +4 is the result of an outstanding shares change on day +4.

[c] These are residuals from the structural portion of a first-order autoregressive model on market volume.

[d] These are standardized residuals from the structural portion of a first-order autoregressive model on market volume.

[e] These are residuals from the structural portion of a first-order autoregressive model on total number of NYSE transactions.

[f] These are standardized residuals from the structural portion of a first-order autoregressive model on total number of NYSE transactions.

the *UTV030* measure is significant at the ten percent level on days −8, −6, and +1; at the five percent level on day 0, day +3, day +4, day +6, and day +7; and at the one percent level on day +5. It is also significant at the ten percent level for the day −2 to +2 time period, and at the five percent level for the day −4 to +5 and −9 to +10 periods. Thus, we reject hypothesis H2.3 and for individual investors conclude that a trading response occurs around annual report release dates.

Table 7
Mean Unexpected Numbers of Transactions Stratified by Value at Annual Report Dates

Day(s) Relative to Report Date[a]	*Number of Reports*[b]	*Mean Actual Less Expected Numbers of Transactions:*[c]			
		Transaction Size <$30,000		*Transaction Size >$100.000*	
		Unstand.	*Stand.*	*Unstand.*	*Stand.*
−9	535	0.028	0.008	−0.169	−0.044
−8	541	0.523	0.061†	−0.137	0.028
−7	547	−0.824	−0.036	0.131	0.032
−6	555	−0.160	0.064†	−0.146	−0.058
−5	563	−0.748	−0.018	0.026	−0.022
−4	563	−0.509	0.023	−0.011	−0.023
−3	563	−0.572	−0.016	−0.067	0.001
−2	562	0.102	0.059	0.215	0.004
−1	563	0.001	−0.009	0.637	0.069†
0	563	0.304	0.085‡	0.417	0.019
+1	563	−0.113	0.064†	0.156	0.030
+2	563	−0.432	0.047	0.143	0.010
+3	563	0.469	0.084‡	0.404	0.040
+4	562	0.844	0.098‡	0.268	−0.003
+5	562	1.120	0.139*	0.031	0.015
+6	562	0.951	0.112‡	0.212	0.054
+7	547	0.419	0.081‡	0.123	0.034
+8	527	−0.336	0.046	0.070	−0.007
+9	519	−0.396	0.005	−0.073	−0.010
+10	505	0.165	0.061	−0.022	−0.032
−2–+2	563	0.088	0.246†	1.567	0.131
−4–+5	563	1.437	0.574‡	2.192	0.160
−9–+10	478	2.694	1.116‡	2.660	0.151

† Significant at the ten percent level (one-tailed cross-sectional *t*-test).
‡ Significant at the five percent level (one-tailed cross-sectional *t*-test).
* Significant at the one percent level (one-tailed cross-sectional *t*-test).

[a] The annual report date is the earlier of the SEC or Disclosure, Inc., stamp date.

[b] Days within ten days of the preliminary announcement date or within five days of the first quarter announcement date are deleted, resulting in the reduced sample sizes on days −9 through −6 and +7 through +10. The loss of one firm on day −2 is because of missing trading data for that firm on day −2. The loss of one firm for days +4 to +10 is the result of an outstanding shares change on day +4.

[c] For Unstand. these are residuals from the structural portion of a first-order autoregressive model on total NYSE transactions. For Stand. these are standardized residuals from the same model.

We have two explanations for the significance of mean number of transactions (*UTR*) and the insignificance of mean volume (*UV*) reported in table 6. First, as noted previously, Cready and Ramanan (1990a) find that tests employing number of transactions are more powerful than those employing volume. Since the trading response appears rather modest (e.g., an average of 5.38 additional transactions per report over days −9 to +10 per the unstandardized design), it is not surprising that a response is detected only by tests employing the more sensitive measure.[24] Second, number of transactions is more sensitive than volume to number of traders. As table 7 demonstrates, much of the report-induced trading occurs within the smaller transaction size classes (i.e., among individual investors). Thus, it is again not surprising that a response is detected by the small-transaction-sensitive number of transactions measure.

---

[24] These unstandardized results should be interpreted cautiously since they are dominated by the set of heavily traded firms within the sample and do not reflect the linearity-correcting influence to trading achieved by using log transformations.

Relative Trading Activity

The tables 6 and 7 results together suggest a trading response occurring over the day −2 to +7 time period and concentrated in the day +3 to +7 period. Accordingly, in examining relative trading activity, we focus on these two multiday periods rather than the more general multiday periods employed in tables 6 and 7.

Table 8 reports share-size-stratified trading for transactions of 2,000 or more shares (i.e., institutional transactions), transactions of 900 or fewer shares (i.e., individual transactions), transactions of between 300 and 900 shares, and transactions of 100 or 200 shares. The 2,000 shares or more and 100 to 900 share results parallel those reported in table 7, indicating a significant trading response among individual investors and an absence of such a response among institutions. Within the individual investor (100 to 900 share) share size stratum, the mean unexpected number of 100 and 200 share transactions is significantly positive (at the ten percent level or better) on each day, except day +8, in the day 0 to +10 time period and in both multiday periods. In contrast, the mean unexpected number of transactions in the 300 to 900 share stratum is significant at the ten percent level on day +3, day +8, the day +3 to +7 period inclusive, the day −2 to +7 period inclusive, and at the five percent level on only day +1. Viewing transaction size as a proxy for investor wealth, these results suggest that the individual investor trading response is greater among less wealthy individual investors.

While the table 8 results suggest that those who trade in response to the annual report are primarily less wealthy (small) individual investors, the results should be interpreted cautiously. The likelihood of detecting a given percentage increase in number of transactions (at a given level of statistical significance) almost certainly varies across strata. Thus, an alternative explanation is that the likelihood of detecting a given percentage increase in trading increases as transaction size stratum decreases.

Mean transaction size (i.e., *MTR* and *MTR9*) results are reported in table 9. *MTR* is frequently less than expected in the time period surrounding the annual report date, including both multiday periods examined. The unstandardized difference between actual and expected *MTR* is significant only on day +4 (ten percent level), while the standardized difference is significant at the five percent level in both multiday periods. While the multiday standardized results indicate that individual trading is heavier than institutional trading, the unstandardized results (upon which our hypotheses are constructed) do not support such a conclusion. Overall, we view the *MTR* results as inconclusive. *MTR9*, on the other hand, is significantly less than expected in the day +3 to +7 time period (one percent level) and in the day −2 to +7 time period (five percent level) regardless of standardization. It is also significantly less than expected on days +5, +6, and +7. Thus, small transaction size investors appear to dominate the individual investor annual report trading response. Furthermore, the occurrence of significant negative differences in days +5, +6, and +7, but not on earlier days, is consistent with the proposition that the subset of wealthier investors who acquire the information do so more quickly than less wealthy investors, which is also reported in Cready (1988).

Table 8
Mean Unexpected Numbers of Transactions Stratified by Share Size at Annual Report Dates

		Mean Actual Less Expected Numbers of Transactions:[c]							
		Transactions of 2,000 or More Shares		*Transactions of 900 or Fewer Shares*		*Transactions of 300–900 Shares*		*Transactions of 100 and 200 Shares*	
Day(s) Relative to Report Date[a]	*Number of Reports*[b]	*Unstand.*	*Stand.*	*Unstand.*	*Stand.*	*Unstand.*	*Stand.*	*Unstand.*	*Stand.*
−9	402	−0.204	−0.017	−0.232	−0.003	−0.019	0.0270	−0.200	−0.0154
−8	405	0.031	0.021	0.324	0.035	0.071	0.0011	0.260	0.0434
−7	411	0.226	0.013	−0.867	−0.045	−0.402	−0.0769	−0.478	−0.0239
−6	416	0.177	0.021	0.026	0.067	−0.181	0.0378	0.215	0.0867‡
−5	421	0.035	−0.001	−0.945	−0.047	−0.725	−0.0772	−0.214	0.0037
−4	422	−0.104	−0.018	−0.417	0.045	−0.110	0.0379	−0.305	0.0503
−3	422	−0.071	0.015	−0.475	0.023	−0.293	0.0009	−0.182	0.0433
−2	421	0.171	0.009	−0.346	0.014	−0.200	−0.0081	−0.147	0.0205
−1	422	0.082	0.015	−0.257	−0.031	−0.127	−0.0251	−0.130	−0.0245
0	422	0.082	0.020	−0.192	0.099‡	−0.380	0.0259	0.202	0.1013‡
+1	422	0.076	0.037	0.440	0.117‡	0.266	0.0918‡	0.193	0.0704†
+2	422	0.050	0.009	−0.229	0.058	−0.179	0.0449	−0.039	0.0674†
+3	422	0.347	0.037	0.334	0.089‡	−0.004	0.0676†	0.345	0.0689†
+4	421	0.306	0.061	1.010	0.134‡	0.000	0.0485	1.012	0.1517*
+5	421	−0.006	0.049	1.320	0.166*	0.144	0.0466	1.174	0.1947*
+6	421	0.168	0.006	1.040	0.134‡	0.366	0.0578	0.673	0.1386*
+7	409	0.093	−0.004	0.723	0.103‡	0.239	0.0439	0.484	0.1129‡
+8	394	−0.068	−0.011	−0.164	0.053	−0.031	0.0687†	−0.113	0.0340
+9	389	0.052	−0.022	0.133	0.025	−0.227	−0.0450	−0.386	0.0705†
+10	378	0.072	−0.005	0.776	0.115‡	0.188	0.0550	0.617	0.1213‡
+3–+7	410	0.996	0.167	4.171	0.627*	0.702	0.256†	3.473	0.669*
−2–+7	410	1.507	0.239	3.277	0.850*	−0.008	0.371†	3.334	0.872*

† Significant at the ten percent level (one-tailed cross-sectional *t*-test).

‡ Significant at the five percent level (one-tailed cross-sectional *t*-test).

* Significant at the one percent level (one-tailed cross-sectional *t*-test).

[a] The annual report date is the earlier of the SEC or Disclosure, Inc., stamp date.

[b] Days within ten days of the preliminary announcement date or within five days of the first quarter announcement date are deleted, resulting in the reduced sample sizes on days −9 through −6 and +7 through +10. The loss of one firm on day −2 is because of missing trading data for that firm on day −2. The loss of one firm for days +4 to +10 is the result of an outstanding shares change on day +4.

[c] For Unstand. these are residuals from the structural portion of a first-order autoregressive model on total NYSE transactions. For Stand. these are standardized residuals from the same model.

Table 9
Unexpected Mean Transaction Sizes in Shares at Annual Report Dates

		Mean Actual Less Expected Mean Transaction Size in Shares:			
		All Transactions[c]		*Transactions of 900 or Fewer Shares*[d]	
Day(s) Relative to Report Date[a]	*Number of Reports*[b]	*Unstand.*	*Stand.*	*Unstand.*	*Stand.*
−9	402	37.9	0.031	4.01	0.0630
−8	406	72.9	0.072	3.34	0.0107
−7	411	35.3	0.071	−1.86	−0.0329
−6	416	−50.1	−0.027	−0.89	−0.0382
−5	421	−0.2	−0.009	−4.59	−0.0890†
−4	422	−20.9	−0.043	−0.55	0.0279
−3	422	24.3	−0.009	−2.96	−0.0347
−2	421	27.6	−0.006	0.32	0.0295
−1	422	−82.4	−0.043	−0.02	0.0366
0	422	−57.3	−0.027	−2.81	−0.0841
+1	422	−57.0	−0.040	1.91	0.0299
+2	422	46.7	−0.008	−3.33	−0.0715
+3	422	22.9	−0.040	1.30	−0.0297
+4	421	−88.9†	−0.055	−2.18	−0.0745
+5	421	−76.2	−0.058	−5.15	−0.1108‡
+6	421	−6.9	−0.048	−5.99‡	−0.0804†
+7	409	2.1	−0.022	−5.25†	−0.0907‡
+8	394	−4.0	−0.017	−1.04	−0.0228
+9	388	26.1	−0.004	−3.65	−0.0838
+10	378	24.9	0.003	−4.17	−0.0859†
+3–+7	410	−43.6	−0.273‡	−3.77*	−0.403*
−2–+7	410	−34.1	−0.377‡	−2.19‡	−0.441‡

† Significant at the ten percent level (two-tailed cross-sectional *t*-test).
‡ Significant at the five percent level (two-tailed cross-sectional *t*-test).
* Significant at the one percent level (two-tailed cross-sectional *t*-test).

[a] The annual report date is the earlier of the SEC or Disclosure, Inc., stamp date.

[b] Days within ten days of the preliminary announcement date or within five days of the first quarter announcement date are deleted, resulting in the reduced sample sizes on days −9 through −6 and +7 through +10. The loss of one firm on day −2 is because of missing trading data for that firm on day −2. The loss of one firm for days +4 to +10 is the result of an outstanding shares change on day +4.

[c] For Unstand. these are residuals from a time-series regression where daily mean transaction size in shares calculated over all transactions and is the dependent variable and trading day (per CRSP) is the independent variable. For Stand. these residuals are further divided by the regressions' standard error.

[d] For Unstand. these are residuals from a time-series regression where daily mean transaction size in shares calculated over transactions between 100 and 900 shares is the dependent variable and trading day (per CRSP) is the independent variable. For Stand. these residuals are further divided by the regression's standard error.

IV. Conclusion

This analysis documents an increase in the number of transactions, weaker evidence of an increase in volume of shares traded, and the absence of any sizable price response at annual report dates. We interpret this as evidence that: (1) at their release, annual reports are informative to investors in making investment decisions; and (2) that annual reports appear to provide a social benefit to investors since the condition of price invariance appears to hold.

The assumption of price invariance, however, can be debated. Other studies present findings favoring both sides of the price invariance issue. We do, however, conclude via simulation that any price change accompanying the annual report is on average small (i.e., less than 0.75 percent) in terms of an observable one-day effect in the time period immediately surrounding the annual report release date.[25] We further believe that a hypothetical price change of such magnitude lacks much economic significance in terms of addressing social welfare questions within the Lev and Ohlson (1982) framework.

Our failure to detect much of a volume response along with an estimated increase in the number of transactions (table 6) of between four and five transactions per report may suggest that only a very small number of investors find annual reports informative. Such a conclusion is unwarranted, however, since trading reflects the actions of only those investors who use the information and, as a result, transact; a far greater number of investors make use of the annual report but take no immediate trading action.

Perhaps most interesting is observing that the trading response is greatest among less wealthy individual investors, as suggested by Hakansson (1977), while other wealthier and more sophisticated investors rely on alternative predisclosure information sources in making investment decisions. It also implies, unlike Cready (1988), that in a multidisclosure environment, information value does not necessarily increase with wealth.

Finally, the trading response period to annual reports appears relatively long, extending to day +7 relative to the report date, and suggests a rather lengthy information assimilation process on the part of many investors with respect to annual reports. Indeed, the heaviest portion of the response appears to occur in the neighborhood of days +5 and +6, as opposed to the more typical day −1 and 0 (e.g., trading in response to preliminary earnings announcements (Morse 1981)).

---

[25] In further simulation results not reported here, only modest reductions in rejection probabilities occur when the price response is spread over three-day periods. Of course it may well be that weeks or even months are required before one observes the full annual report price effect. Such a hypothetical long-term price effect is an empirical question relevant to this paper's conclusions, but well beyond its scope.

Appendix A
Description of Returns Simulation Procedure

The simulation procedure is similar to that of Brown and Warner (1980) in that we constructed 250 portfolios, each consisting of 563 firm-dates randomly selected from the population of NYSE-listed and CRSP returns-available firms over years 1981 through 1984. Mean event period *SUR* and *ABUR* are calculated for each portfolio in the manner described in the paper. Day 0 excess returns are artificially induced by adding a random variate to the day 0 returns of each portfolio observation. Each variate is drawn from a normal distribution with mean 0 and standard deviation chosen to correspond to a specified mean absolute return effect. For example, the expected absolute value of a variable distributed $N(0, 0.0125)$ equals 0.01, and the 0.01 inducement is achieved by generating an $\tilde{x}$, $\tilde{x} \sim N(0, 0.0125)$, for each firm-date in each portfolio and adding it to the existing day 0 return for that firm-date. Other expected absolute values examined and their normal distribution equivalents are 0.005, random variable distributed $N(0, 0.00625)$, and 0.0075, random variable distributed $N(0, 0.009375)$.

For each level of return inducement we determined percentages of portfolios with *SUR*- and *ABUR*-based P-levels of less than 0.5, less than 0.3, and less than 0.2 for one-day, three-day, and five-day "event" windows. Table A1 provides a comprehensive summary of these percentages.

Of note in table A1 is that when no return inducement is present the rejection frequencies for *SUR* are much higher than the P-level, meaning that the unadjusted *SUR* metric is biased in favor of the alternative hypothesis. This is consistent with the *SUR* simulation findings reported in chapter 5 of Marais (1985). The rejection frequencies for *ABUR*, while generally higher than the P-level, are not so much higher as to raise major concerns regarding overrejection. Of greater interest, the rejection frequencies for *ABUR* generally exceed those of *SUR* in the presence of a non-zero induced return, confirming the earlier proposition from Rohrbach and Chandra (1989) that absolute value return metrics outperform squared return metrics in detecting excess returns.

Table A1
Rejection Frequencies of the Null Hypothesis of No Excess Return for 250 Portfolios of 563 Firms Each

			Induced One-Day Return Percentage (Standard Deviation of Induced Disturbance Term)			
P-Level	*Event Period*	*Metric*	*0.0 (0.0)*	*0.5 (0.00625)*	*0.75 (0.009375)*	*1.00 (0.0125)*
0.5	1 DAY	*SUR*	61.2	91.6	100.0	100.0
		ABUR	55.6	98.0	100.0	100.0
	3 DAY	*SUR*	69.2	85.2	98.4	100.0
		ABUR	55.6	88.4	99.6	100.0
	5 DAY	*SUR*	72.4	84.0	93.6	99.6
		ABUR	55.2	78.8	96.4	100.0
0.3	1 DAY	*SUR*	40.0	81.6	99.2	100.0
		ABUR	37.2	94.4	100.0	100.0
	3 DAY	*SUR*	48.8	74.0	90.8	99.6
		ABUR	30.8	74.8	96.4	100.0
	5 DAY	*SUR*	51.2	68.0	86.4	96.4
		ABUR	33.2	61.6	88.4	99.6
0.2	1 DAY	*SUR*	24.4	75.2	98.8	100.0
		ABUR	23.6	92.4	100.0	100.0
	3 DAY	*SUR*	30.8	60.0	84.4	98.8
		ABUR	21.6	64.0	94.0	100.0
	5 DAY	*SUR*	36.4	52.4	77.6	92.4
		ABUR	22.4	51.2	78.8	99.2

References

Ajinkya, B. B., and P. C. Jain. 1989. The behavior of daily stock market trading volume. *Journal of Accounting & Economics* 11 (November): 331–60.

Bamber, L. S. 1986. The information content of annual earnings releases: A trading volume approach. *Journal of Accounting Research* 24 (Spring): 40–56.

———. 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *The Accounting Review* 62 (July): 510–32.

Beaver, W. H. 1968. The information content of annual earnings announcements. *Journal of Accounting Research* 6 (Supplement): 67–92.

Bernard, V. L., and T. L. Stober. 1989. The nature and amount of information in cash flows and accruals. *The Accounting Review* 64 (October): 624–52.

Brown, S. J., and J. B. Warner. 1985. Using daily stock returns: The case of event studies. *Journal of Financial Economics* 14 (March): 3–31.

Cready, W. M. 1988. Information value and investor wealth: The case of earnings announcements. *Journal of Accounting Research* 26 (Spring): 1–27.

———, and R. Ramanan. 1990a. Detecting trading responses using transactions-based research designs: A simulation analysis. Working paper, University of North Carolina at Chapel Hill (August).

———, and ———. 1990b. The power of tests employing log-transformed volume in detecting abnormal trading. Working paper, University of North Carolina at Chapel Hill (April).

Foster, T. W., III, D. R. Jenkins, and D. W. Vickery. 1986. The incremental information content of the annual report. *Accounting and Business Research* 16 (Spring): 91–98.

Hakansson, N. H.. 1977. Interim disclosure and public forecasts: An economic analysis and a framework for choice. *The Accounting Review* 52 (April): 396–416.

———, J. G. Kunkel, and J. A. Ohlson. 1984. A comment on Verrecchia's no trading theorem. *Journal of Accounting Research* 22 (Autumn): 765–67.

Howard, P. F. 1982. Transactions, volume, and value as measures of investors' response to information. Working paper, Monash University (November).

Jain, P. 1986. Analysis of the distribution of security market model prediction errors for daily returns data. *Journal of Accounting Research* 24 (Spring): 94–95.

Jang, J., and B. T. Ro. 1989. Trading volume theories and their implications for empirical information content studies. *Contemporary Accounting Research* 6 (Fall): 242–61.

Lev, B. 1988. Toward a theory of equitable and efficient accounting policy. *The Accounting Review* 63 (January): 1–22.

———, and J. Ohlson. 1982. Market-based empirical research in accounting: A review, interpretation, and extension. *Journal of Accounting Research* 20 (Supplement): 249–322.

Lobo, G. J., and I. Song. 1989. Historical cost income and its cash and accrual components. *The Accounting Review* 64 (April): 329–43.

Marais, M. L. 1985. Some applications of computer intensive statistical methods to empirical research in accounting. Ph.D. dissertation, Stanford University.

Morse, D. C. 1981. Price and trading volume reactions surrounding earnings announcements: A closer examination. *Journal of Accounting Research* 19 (Autumn): 374–83.

Mynatt, P. G. 1988. The information content of financial statement releases. Ph.D. dissertation, University of North Carolina at Chapel Hill.

Patell, J. M. 1976. Corporate forecasts of earnings per share and stock price behavior: Empirical tests. *Journal of Accounting Research* 14 (Autumn): 246–76.

Pincus, M. 1983. Information characteristics of earnings announcements and stock market behavior. *Journal of Accounting Research* 21 (Spring): 166–83.

Richardson, G., S. Sefcik, and R. Thompson. 1986. A test of dividend irrelevance using volume reactions to a change in dividend policy. *Journal of Financial Economics* 15 (December): 313–33.

Rohrbach, K., and R. Chandra. 1989. The power of Beaver's U against a variance increase in market model residuals. *Journal of Accounting Research* 27 (Spring): 145–55.

Shareownership 1983. 1984. The New York Stock Exchange, New York, April.

Verrecchia, R. E. 1981. On the relationship between volume and consensus of investors: Implications for interpreting tests of information content. *Journal of Accounting Research* 19 (Spring): 271–83.

Wilson, G. P. 1987. The incremental information content of the accrual and funds components of earnings after controlling for earnings. *The Accounting Review* 62 (April): 293–322.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 313–332

Auditor Credibility and Initial Public Offerings

Krishnagopal Menon
Boston University
David D. Williams
The Ohio State University

SYNOPSIS: An important differentiating attribute of the audit product is believed to be the credibility that the auditor is perceived to bring to an audit engagement. This study uses the context of the initial public offering (IPO) to investigate auditor credibility.

It is contended that information asymmetry problems lead to a demand for credible auditors in companies going public. Entrepreneurs have incentives to signal their knowledge of favorable future earnings by selecting reputable auditors. Since there is limited information available on firms going public, employing credible auditors can convey monitoring cost advantages as well. Investment bankers also have a preference for credible auditors since they rely on audited financial statements in certifying the value of the firm and determining whether to underwrite the offering.

In the present study, we consider auditor credibility in IPOs from the perspective of the client and the investment banker. If there is an increased demand for auditor credibility at the time of the IPO, there should be a significant number of credibility-increasing auditor changes prior to the offering. Further, if the investment banker benefits from having a more credible auditor sign off on statements prepared by the client, this should be reflected in the investment banker's fee structure.

The empirical analysis is performed on companies that went public in 1985 and 1986. Relatively few auditor changes are observed prior to the offering. However, among those companies making auditor changes, there is a clear preference for more credible auditors. Logistic regression analysis shows that companies with prestigious investment bankers are more likely to change away from local auditors to more credible CPAs. The type of

We are grateful to Jane Mutchler, Steve Kaplan, and two anonymous referees for helpful comments on earlier versions of this paper. This research was supported in part by grants from the Institute for Accounting Research and Education at Boston University and the Peat Marwick Foundation.

Submitted April 1989.
Revisions received August and October 1990.
Accepted November 1990.

underwriting arrangement employed is also significant, consistent with an investment banker preference for credible auditors.

A regression analysis is conducted, using the 1985 and 1986 IPOs, modeling investment banker compensation as a function of several factors, including type of auditor employed by the issuing firm. In the case of "firm commitment" offerings, the auditor type is found to be significant. Clients seem to be charged a smaller investment banking fee if they are associated with Big Eight auditors. There is no apparent auditor effect in the case of "best efforts" offerings. The evidence generally supports the hypothesis that investment bankers and their clients have a preference for credible auditors for the IPO.

Key Words: *Auditor credibility; Initial public offerings; Auditor changes; Investment banker compensation.*

Data Availability: *The data used in this study are available from the authors upon request.*

THE paper is organized as follows. Section I discusses the relevant theoretical issues related to auditor credibility in companies making IPOs and derives testable implications. Section II describes the samples used and the methods employed to analyze the data. Results are also presented in this section. The final section contains a summary of the study and some concluding remarks.

I. IPOs, Investment Bankers, and Auditor Choice

Auditor credibility is characterized in recent research as one attribute of a differentiated audit product (DeAngelo 1981a, 1981b; Simunic and Stein 1987).[1] It is contended that the credibility of financial statements, in part, depends on the perceived quality of the audit. Since the actual quality of the audit is not observable, auditors attempt to communicate their quality through such signals as brand names and reputation. As a consequence, different levels of credibility are offered in the audit market at different prices (Simunic and Stein 1987).

The reputation of the auditor is particularly likely to affect the perceived credibility of financial statements when a firm makes its initial public offering (IPO) of stock. Companies making IPOs typically are little known to investors, who, in the absence of alternative sources of information, must place substantial reliance on management's reports. If financial statement users differentiate between levels of auditor credibility, this should be more readily detected in the IPO market than in markets for older, better-known companies.

Anecdotal evidence suggests that companies going public do consider the credibility offered by different CPA firms while selecting an auditor. The emphasis on selecting a reputable auditor is clearly conveyed in the following advice, provided in a manual for entrepreneurs planning to make an IPO.

---

[1] By auditor credibility we mean perceived auditor quality. DeAngelo (1981b) defines auditor quality as "the market assessed joint probability that a given auditor will both (a) discover a breach in the client's accounting system, and (b) report the breach." We prefer to use the term *auditor credibility* rather than *auditor quality* because it is the perception of external users of financial statements that is important in this determination. We refer to perceived high-quality auditors as "credible" or "reputable."

> One key consideration is which accounting firm to hire. Many investment bankers insist that one of the "Big Eight" accounting firms be used. While this may not seem necessary, to the underwriter it is. The major accountants have been through the process many times before and are less likely to miss some important aspect of reporting. Perhaps the most important aspect, however, is that of image. A financial statement audited by a reputable, well-known accounting firm carries more weight than an unknown one (Sutton and Benedetto 1988).

In this paper, we refer to the belief that investment bankers and their clients have a preference for credible auditors when making an IPO as the auditor credibility hypothesis.

Prior researchers have studied the valuation effects of selecting a credible auditor for the IPO. Simunic and Stein (1987) examine the market premium over book value (the ratio of the market value of equity to its book value, computed subsequent to the issue) for a sample of firms making IPOs. They find that firms associated with Big Eight rather than non-Big Eight auditors are able to obtain a higher premium.

Following Beatty and Ritter (1986) and Rock (1986), Beatty (1989) and Balvers et al. (1988) contend that underpricing, or the difference between the offering price of the IPO and the first day closing market price, is a function of the *ex ante* uncertainty about the value of the issue. Engaging a reputable auditor helps to reduce this uncertainty. Beatty's results suggest that underpricing is lower for firms with higher-priced (and presumably more reputable) auditors. Balvers et al. also find that the clients with more reputable auditors have lower underpricing, though the magnitude of the effect of the auditor's reputation decreases as the investment banker's reputation increases.

The finding of Simunic and Stein (1987), Beatty (1989), and Balvers et al. (1988) that IPO market participants are willing to pay a premium for auditor credibility is consistent with the auditor credibility hypothesis. This hypothesis has two testable implications. First, there should be significant pre-IPO auditor switching activity as companies replace small audit firms with more prestigious national or Big Eight auditors.[2] Second, investment bankers should charge higher fees to companies associated with less credible auditors. We provide empirical evidence on these two implications in this study. In the remainder of this section, we discuss theoretical issues related to incentives of clients and their investment bankers to select credible auditors for the IPO.

Client Incentives

When firms make IPOs, the issuer has private information about the firm's future prospects that is unavailable to prospective investors. This information asymmetry problem may lead to a demand for auditor credibility for two reasons: first, to signal firm value, and second, to minimize monitoring costs.

According to the signaling argument, in order to avoid an adverse selection problem, entrepreneurs signal their private information about future earnings. One available signal is the choice of auditor (Titman and Trueman 1986; Bachar 1989). Titman and Trueman (1986) show analytically that an entrepreneur has the incentive to choose the level of auditor quality that correctly reveals the entrepreneur's private information about the firm. It is not worthwhile for an entrepreneur with unfavorable information about the firm to select a more credible auditor. The more reputable auditors charge a premium price for the additional credibility they bring. Also, such auditors are more likely to uncover and disclose adverse information about the firm.

---

[2] We use the term "Big Eight" because there were eight firms in this class of auditors during the period covered by the study. Mergers have since reduced the number of firms in this class to six.

The monitoring costs argument is based on the limited information available to prospective investors in an IPO to assess the validity of managers' claims. IPOs tend to be made by small firms with limited operating histories. There is generally little information available on these firms at the time of the offering. Investors must rely heavily on the content of the entrepreneur/manager's self disclosures in order to evaluate the performance and future prospects of the firm. In the absence of credible financial statements, market participants have to undertake a costly search for alternative sources of information about the firm.[3] Financial statement credibility, which is enhanced by association with a reputable auditor, reduces monitoring costs.

Investment Banker Incentives

The investment banker (underwriter) has incentives to induce the client firm to select a credible auditor. Anecdotal evidence suggests that the client's decision to replace a small auditor with a prestigious one is influenced frequently by the investment banker (Balvers et al. 1988, 608; Jacobs 1983; Sutton and Benedetto 1988).

The investment banker engages in two types of activities in a public offering: a distribution activity and an underwriting activity (Mandelker and Raviv 1977). In carrying out the distribution activity, the investment banker acts as a financial intermediary, assisting the issuing firm in designing and timing the offering and in distributing the securities. The underwriting function entails providing a guarantee to the issuer against the risk of fluctuations in the price of the offered securities. The two common offering arrangements are "firm commitment" (FC) and "best efforts" (BE). In a FC contract, the investment banker assumes both underwriting and distribution functions. The banker buys the offered securities for less than the offering price and resells them to investors. In a BE arrangement, the investment banker promises to sell as many of the offered securities as possible but takes no responsibility for unsold securities.

The two different investment banking functions lead to two distinct types of risk. In underwriting securities, the investment banker accepts the risk of not being able to resell all of the offered securities and having to retain the unsold remainder. However, distribution carries risk as well. Since the issuing entrepreneur has private information about the firm's economic value not known to potential investors, one reason the entrepreneur engages the investment banker is to attest to the firm's value. The investment banker certifies that the offering price reflects the available information about the firm, both public and private. The banker suffers from a loss of reputational capital if the market later concludes that the client provided misleading information about itself or that the offering was inappropriately priced (Beatty and Ritter 1986; Booth and Smith 1986). The FC arrangement carries more risk for the investment banker since it includes both underwriting and distribution functions.

A reputable auditor helps to reduce the investment banker's risk in at least two ways. First, a credible auditor can provide the investment banker with more assurance about the financial numbers upon which the offering price is based and reduce the investment banker's own information search costs. Second, for the signaling and monitoring reasons discussed earlier, an IPO associated with a prestigious auditor is likely to be more attractive to investors.

---

[3] Tinic (1988) notes that the amount of stock that an investor can be expected to be allocated in a popular offering is generally too small to compensate the costs of an adequate information search.

Investment bankers may have other reasons to prefer reputable auditors, unrelated to information asymmetry. Tinic (1988) points out that securities regulations impose significant legal requirements upon investment bankers, accountants, and other professionals involved in public offerings. These professionals are expected to diligently examine and disclose every material fact in relation to an offering. The investment banker can face large legal liabilities if investors make stock purchases on the basis of statements that are deemed to omit relevant facts or to be misleading. Auditors, too, frequently incur liabilities because of their association with IPOs (Hall and Renner 1988).

Tinic (1988) contends that investment bankers seek insurance for their potential liability.[4] It could be argued that the IPO's association with a reputable auditor provides a form of insurance. Not only can the investment banker get greater comfort from relying on statements audited by reputable CPAs, but the most prestigious auditors are perceived as having "deep pockets" of their own, which may mitigate the potential financial loss to the investment banker in the event of litigation.

If investment bankers and clients prefer credible auditors at the time of the IPO, two implications follow. First, the increased demand for auditor credibility may lead to the firm switching auditors prior to the public offering. Second, the compensation demanded from the issuing firm by the investment banker should be adjusted for the level of credibility offered by the firm's auditor.

An increased demand for auditor credibility need not result in an auditor change. Firms may select reputable auditors several years prior to the IPO for reasons unrelated to their credibility.[5] For example, small clients may demand reputable auditors' financial, tax, or consulting expertise.[6] Firms also may engage reputable auditors because of a demand for auditor credibility that arises before the public offering, stemming from the need to obtain capital from banks, venture capitalists, or private investors.

In addition, firms must weigh the potential benefits of acquiring a more reputable auditor for the IPO against the increased cost. The signaling benefit is that the auditor's reputation signals the quality of the offering, reducing the uncertainty around the offering price and allowing the offering firm to price the issue closer to its "true" price. The monitoring cost benefit is that the lower cost incurred in monitoring firm performance should result in investors being willing to pay a higher price for the issued stock.

On the cost side, auditors who offer high reputation levels are also likely to charge a premium fee (Palmrose 1986; Francis and Simon 1987). The offering firm must compare the increased economic benefits from employing a more reputable auditor for the IPO with the premium paid for the added credibility in determining whether to change auditors. Further, some firms may view the signal sent by an auditor switch prior to an

---

[4] Tinic (1988) contends that the underpricing phenomenon in IPOs can be attributed to investment bankers lowering the price of the security as a form of implicit insurance against potential future liabilities related to the offering. Tinic's empirical results provide some support for this argument.

[5] Simunic and Stein (1987) classify the demand for audit product attributes into three types: control (the reduction of internal agency costs), product line (the availability of nonaudit services), and credibility.

[6] Many Big Eight and national firms have established "small business" or "emerging business" departments in an attempt to attract young, growing companies, perhaps many years prior to the IPO (Berton 1986).

IPO as conveying costs rather than benefits, fearing that investors may read the change as a red flag that indicates either a dispute with the predecessor auditor or an attempt to suppress information.

Some empirical evidence of auditor switching near the time of the IPO is provided by a questionnaire survey conducted by Carpenter and Strawser (1971). As many as 91 of the 162 respondents (56 percent) in Carpenter and Strawser's sample of IPO companies had retained their present auditor for fewer than five years. However, the study does not provide any direct evidence on whether small auditors are displaced by larger, more credible ones, since there is no information on the predecessor auditor. Neither is it clear whether the auditor change took place before or after the IPO.[7]

In this study, we examine auditor choices made by firms for the IPO and analyze the auditor-switching activity. We use a logistic regression analysis to identify characteristics of firms that replace their small auditors with larger ones, to see if these changes are consistent with a demand for higher credibility. We also assess the effect of auditor credibility on investment banking costs to see if the fees charged to clients making IPOs are reduced as the credibility of the auditor increases.

II. Method and Results

The sample for the study consists of all companies that filed a final prospectus with the Securities and Exchange Commission (SEC) in 1985 and 1986 for their initial public offering of common stock. These firms are identified from the Registered Offerings Statistics (ROS) tape maintained by the SEC.[8] In addition to identifying the names of the firms, the ROS tape provides information on the number of shares issued, the offering price, underwriter commissions, and the type of issue, i.e., whether the underwriting arrangement was FC (800 firms) or BE (520 firms).[9]

Auditor Changes

Two separate analyses related to auditor choices around the time of the IPO are conducted. First, we examine the frequency and nature of auditor changes prior to the IPO. Next, we use a logistic regression model to identify characteristics of firms seeking a higher level of auditor credibility.

Information on the incumbent auditor at the time of the IPO is obtained primarily from the final prospectus. Where the prospectus is unavailable, we rely on the initial filings made at the time of the IPO or the first 10-K filed after the public offering, cor-

---

[7] While Carpenter and Strawser (1971) provide interesting findings, their study is unclear on a couple of points. First, their questionnaire identifies companies that had retained their auditor for less than five years, rather than companies that changed auditors. While the study notes that only very few firms from the sample had been in existence for fewer than five years, it is plausible that a large number of the firms, while having been in existence for at least five years, had only recently employed their first auditor. Second, it is not clear from the description of the procedures followed in the study that all the auditor changes took place prior to the IPO. It is possible that many companies changed auditors shortly after the public offering, but prior to responding to the questionnaire. The large number of auditor changes subsequent to the IPO that we detect in our sample makes this a likely scenario.

[8] Certain financial institutions that went public in 1985 and 1986 are not included in the ROS tape because they filed their offerings with other regulatory organizations (e.g., the Federal Home Loan Bank Board) rather than with the SEC.

[9] The ROS tape contains several errors. In the majority of cases we cross-check data obtained from the ROS tape with the prospectus or other sources.

Table 1
Auditor Selection*

	IPO Sample			
Auditor	Firm Commitment (%)	Best Efforts (%)	Total (%)	Population of Public Companies (%)
Big Eight	574 (76.2)	100 (24.1)	674 (57.7)	6,844 (68.8)
National	88 (11.7)	53 (12.8)	141 (12.1)	1,001 (10.1)
Local	91 (12.1)	262 (63.1)	353 (30.2)	2,106 (21.1)
	753 (100.0)	415 (100.0)	1,168 (100.0)	9,951 (100.0)
No data	47	105	152	
Total	800	520	1,320	9,951

* Population numbers are as of the end of 1986, based on companies covered by *Who Audits America.*

roborated by evidence from secondary sources such as *Going Public: The IPO Reporter*, *Who Audits America*, and the *Public Accounting Report* (*PAR*) list of auditors of IPOs. Out of the 1,320 firms in the sample, we are able to identify the auditor of record at the time of the IPO in 1,168 cases (753 FC and 415 BE).

Auditors are classified into three categories: Big Eight, National, and Local.[10] This classification is used to represent auditor credibility in this study. In reality, auditor credibility is probably a continuum. Although many prior studies employ a simple Big Eight-Other dichotomy, we prefer to use three classes. The Big Eight are presumed to be the most credible, followed by other national auditors and then local firms.

Table 1 shows the auditor preferences of the firms going public. As indicated in the table, the majority of FC offerings (76.2 percent) are associated with Big Eight auditors, while the majority of the BE clients (63.1 percent) have local auditors. A chi-square test shows that proportions of firms associated with the three types of auditors are different for the FC and BE samples (chi-square$=$356.95 with 2 degrees of freedom, $p<0.001$). Simunic and Stein (1987) report a similar association between the type of underwriting arrangement and the auditor, using their data from 1981 IPOs.[11]

We use the 9,951 publicly traded companies covered by *Who Audits America* as of the end of 1986 as a basis for comparison. Of these firms, 68.8 percent are audited by the Big Eight, 10.1 percent by national firms, and 21.1 percent by local firms. In contrast, the frequency of locally audited firms in the BE sample appears to be relatively high, while the frequency of locally audited FC companies (12.1 percent) is low. In both BE and FC cases, the proportions of firms using the three classes of auditors are signifi-

---

[10] Eight firms are treated as having been in the "National" category for the purposes of this study (KMG Main Hurdman; Laventhol & Horwath; Grant Thornton; McGladrey, Hendrickson & Pullen; Seidman and Seidman; Pannell, Kerr, Forster; Kenneth Leventhal; and Oppenheim, Appel, Dixon). These are the firms identified by the *Public Accounting Report* as constituting the "second tier" of CPA firms in 1985, on the basis of their revenues (*PAR* 1985).

[11] Simunic and Stein (1987) report a lower proportion of Big Eight auditors for firm commitment offerings (62 percent) than we find (76.2 percent of the cases for which we are able to identify the auditor). They find more Big Eight auditors associated with best efforts offerings (40 percent against our 24.1 percent).

Table 2
Frequency of Auditor Changes*

	Firms in IPO Sample	(%)	*Firms in Population*	(%)
Firms changing auditors	71	(6.4)	889	(8.9)
Firms not changing auditors	1,034	(93.6)	9,062	(91.1)
Firms with complete data	1,105	(100.0)	9,951	(100.0)
Firms with missing data	215			
Total	1,320			

* Auditor changes for the population are based on 1986 changes reported in *Who Audits America*.

cantly different from proportions for the population of public companies (BE sample: chi-square $=428.78$ with 2 degrees of freedom, $p<0.001$; FC sample: chi-square $=35.54$ with 2 degrees of freedom, $p<0.001$).

For the 1,168 firms whose auditor at the time of the IPO is known, we attempt to determine whether an auditor change has been made in the two years prior to the offering. A firm is classified as having changed auditors if it engaged an auditor other than the one associated with the IPO for the fiscal year that ended prior to the public offering or for the preceding year. If a firm did not have any auditor previous to the present one, it is not identified as having changed auditors, even if this auditor was engaged sometime in the two years prior to the IPO. Similarly, if the inception of the firm was less than two years prior to the offering and the firm retained its first auditor for the IPO, it is classified as a nonswitcher.[12] Information on prior auditors is obtained from a number of different sources. The primary source is the consistency phrasing in the auditor's report in the prospectus or 10-K, which typically explicitly acknowledges (or disclaims) prior audits. In many cases we use other sources to confirm the prior auditor, such as pre-IPO audited financial statements, 8-K and 10-K filings, *Who Audits America*, and proxy statements (which sometimes makes reference to the tenure of existing auditors).

We are able to trace the auditor for two prior years for 1,105 firms (table 2). Only 71 of these firms (6.4 percent) changed their auditor in the two years prior to the IPO, suggesting less of a shift in the client firms' relative demand for different audit service than might be expected. The frequency of the changes does not seem high when compared with normal auditor switching activity in public companies. According to the 1987 edition of *Who Audits America*, 889 firms (or 8.9 percent of the 9,951 firms monitored) changed auditors in 1986, a single year, in contrast to the two-year period we allow for the IPO firms.

One factor that reduces the likelihood of finding a significant number of credibility-increasing auditor switches is the large number of firms already employing Big Eight

---

[12] Restricting the time period for analysis to two years prior to the IPO results in overlooking firms that made changes before that time, perhaps in anticipation of going public. However, the longer the time period between the auditor change and a subsequent initial public offering, the more tenuous is the relationship between the two. Data difficulties prevent us from investigating earlier changes.

Table 3
Direction of Auditor Changes

Panel A. Frequency in IPO Sample:

Successor Auditor	Predecessor Auditor Big Eight (%)	National (%)	Local (%)	Total (%)
Big Eight	8 (100.0)	4 (57.1)	39 (69.6)	51 (71.8)
National	0 (0)	2 (0)	7 (12.5)	9 (12.7)
Local	0 (0)	1 (14.3)	10 (17.9)	11 (15.5)
Total	8 (100.0)	7 (100.0)	56 (100.0)	71 (100.0)

*Panel B. Frequency in Public Companies:**

Successor Auditor	Predecessor Auditor Big Eight (%)	National (%)	Local (%)	Total (%)
Big Eight	543 (74.3)	116 (64.8)	140 (39.2)	799 (63.1)
National	82 (11.2)	14 (7.8)	41 (11.5)	137 (10.8)
Local	106 (14.5)	49 (27.4)	176 (49.3)	331 (26.1)
Total	731 (100.0)	179 (100.0)	357 (100.0)	1,267 (100.0)

* Data for auditor switches in the population are for 1985 and 1986, and are drawn from the *Public Accounting Report.*

auditors. As discussed earlier, it is plausible that these auditors are engaged because of credibility or other needs that exist before the IPO. Interestingly, in 56 of the 71 cases of auditor switching, the predecessor auditor is a local CPA (table 3, panel A).

The majority of switches in our sample, 50 out of 71 (70.4 percent), result in more credible auditors being retained for the IPO, while 20 changes (28.2 percent) are to auditors of approximately the same level of credibility, and one (1.4 percent) is to a less credible auditor. In contrast, the *Public Accounting Report* (*PAR* 1986, 3; 1987, 3) indicates that there were 1,267 auditor changes in the 1985–1986 period in the population of all publicly traded companies. Of these, 297 (23.4 percent) were to more credible auditors, 733 (57.9 percent) were in the same auditor class, while 237 (18.7 percent) of the auditor switches were to less credible auditors (table 3, panel B). A chi-square test shows that the proportions of changes belonging to these three categories in the case of the IPO sample are significantly different from the proportions for the *PAR* population (chi-square = 79.06 with 2 degrees of freedom, $p < 0.001$). While the frequency of auditor changes prior to the IPO is not unusually high, there is evidence of small auditor displacement, consistent with Carpenter and Strawser (1971).

The direction of the auditor switches is consistent with the auditor credibility hypothesis. However, one important alternative explanation for the preference for large auditors is that small auditors may not be equipped with the necessary expertise to handle publicly traded clients, with all of the associated regulatory requirements. While this explanation is plausible, we find it unconvincing for the following reason. The fact that many publicly traded clients retain small auditors (both in our sample and in the population of publicly traded firms) suggests that "SEC expertise" is not exclusively possessed by Big Eight firms. If the need for SEC expertise was the dominant reason for switching from local firms, we would have expected to observe more changes to other local and national firms, which can offer the same services at a lower price.[13]

Still, it is possible that audit product attributes other than credibility drove the changes that resulted in more reputable auditors being selected. We use a logistic regression model to see if the firms in our sample making credibility-increasing switches show characteristics that support the auditor credibility hypothesis. The following model is employed.

$$\begin{aligned} AUDCHG = {} & b_0 + b_1(IBANK) + b_2(DILUTE) + b_3(TYPE) \\ & + b_4(SIZE) + b_5(GROWTH) + e, \end{aligned}$$

where:

$AUDCHG$ = a dummy variable that takes the value 0 if the client firm retains the same local auditor and 1 if the client firm replaces the local auditor with a national or Big Eight CPA;
$IBANK$ = investment banker class, rated from 1 to 4, where 1 is the most prestigious;
$DILUTE$ = the percentage of stock ownership dilution caused by the offering;
$TYPE$ = the type of offering, whether "best efforts" (=1) or "firm commitment" (=0);
$SIZE$ = the post-offering size of the firm; and
$GROWTH$ = the increase in the size of the firm.

We expect *IBANK* to be negatively related to the dependent variable. As discussed earlier, investment bankers are expected to have a preference for more prestigious auditors. This preference is likely to be strongest in the case of the most prestigious investment bankers, who have the greatest reputational capital to lose. Clients employing major investment bankers are more likely to make credibility-increasing auditor switches, yielding a negative b_1 coefficient. If b_1 is positive, it indicates that firms are more likely to select perceived high-quality auditors when the investment banker is not prestigious, implying that the choice of investment banker and auditor are potentially substitutable signals of quality.

Monitoring costs should become more important as the percentage of shares retained by the existing investors goes down and the ownership of new investors goes up. Since increased auditor credibility is hypothesized to reduce monitoring costs, there should be a positive relationship between the dependent variable (*AUDCHG*) and *DILUTE*, the percentage of dilution of the existing stockholders' interest. *DILUTE* may also represent the quality of the offering. Leland and Pyle (1977) show that a high level of retained ownership by existing stockholders is a signal of their positive private information to less informed potential stockholders. A negative b_2 coefficient would be consistent with better quality issues preferring more credible auditors.

---

[13] A similar line of reasoning would suggest that since a number of publicly traded companies use non-Big Eight auditors, IPO firms should be able to obtain credibility without engaging a Big Eight auditor. Indeed, any CPA can provide the minimum credibility necessary for an IPO. The contention in this paper is that auditor credibility conveys benefits, and that additional credibility can be acquired at a cost. In many cases, client firms may make the cost-benefit trade-off in favor of non-Big Eight firms.

The variable *TYPE* represents the underwriting arrangement employed. Since the investment banker has more economic incentive in an FC contract than a BE contract, the strength of the investment banker's demand for a credible auditor should correspondingly be greater for FC offerings. If the investment banker influences the client firm's choice of auditor, FC firms should be more likely to make credibility-increasing auditor switches, resulting in a negative coefficient for *TYPE*.

Simunic and Stein (1987) point out that one important characteristic of the audit product is control, which stems from the need to limit the internal agency costs in firms. Demand exists for this audit product attribute even in the absence of external monitoring needs. The level of internal agency costs increases with the size of the firm (Simunic and Stein 1987). To represent the increase in demand for control that accompanies the IPO, we use *GROWTH*, defined as the increase in size. However, *GROWTH* does not reflect the total post-IPO internal agency costs, which may govern the decision to select a larger auditor. To capture this dimension, we use *SIZE* (post-offering size) as an additional variable. Positive coefficients for b_5 and b_6 would support the position that demand for control drives the switching to larger auditors, rather than (or in addition to) a search for greater auditor credibility.

The sample for this analysis is restricted to client firms that had local auditors prior to the IPO. Missing data on nine firms reduced the sample size to 352 firms. The dependent variable is coded as 1 if the client switched to a more credible auditor (46 firms) and as 0 if it stayed in the same auditor credibility class, either retaining its existing auditor (296 firms) or switching to another local auditor (ten firms).[14]

Data on the independent variables are obtained from the ROS tape. The class of investment banker employed for the offering is coded from 1 to 4, where 1 is the most prestigious. Eighteen investment bankers identified as most prestigious by Balvers et al. (1988), based on their consistent appearance in the "Top 25" reported annually by *Institutional Investor*, are classified as "major" and coded as 1. Johnson and Miller (1988) identify an additional set of 16 investment bankers as being prestigious, based on analysis of tombstone advertisements. Thirteen of these bankers are associated with IPOs in our sample. These are classified as "sub-major" and coded as 2. Other investment bankers identified by the ROS tape as "national" or "regional" are coded as 3, while the investment bankers identified by the ROS tape as "state" or "local" are coded as 4. We are able to identify the investment banker for 1,176 IPOs; of these, 251 offerings are associated with "major" investment bankers, 92 with "sub-major," 271 with "national" or "regional," and 562 with "state" or "local."[15]

---

[14] Omitting from the analysis the ten firms that switched from one local auditor to another leads to results practically identical to those reported here.

[15] The "major" investment banker category, coded as 1, consists of: Bear Stearns; Dean Witter Reynolds; Dillon Read; Drexel Burnham Lambert; Donaldson, Lufkin & Jenrette; E. F. Hutton; First Boston; Goldman Sachs; Kidder, Peabody; Lazard Freres; Merrill Lynch; Morgan Stanley; Paine Webber; Prudential-Bache; Rothschild, Unterberg, Towbin; Salomon Brothers; Shearson Lehman; and Smith Barney. The following investment bankers are included in the "sub-major" category, coded as 2: Advest; A. G. Edwards; Alex Brown; Blunt, Ellis; F. Eberstadt; Hambrecht and Quist; Ladenburg, Thalmann; Mosely, Hallgarten; Oppenheimer; Prescott, Ball & Turben; Robinson-Humphrey; Tucker Anthony; and Wheat, First Securities. When the analyses performed in this study using four categories for investment banker type are repeated, combining categories one and two, essentially similar results are obtained.

Table 4
Correlation Matrix for Auditor Choice Model*
(n=352)

	IBANK	DILUTE	TYPE	SIZE	GROWTH
IBANK	1.000 (0.000)	0.196 (0.001)	0.423 (0.001)	−0.572 (0.001)	0.431 (0.001)
DILUTE		1.000 (0.000)	0.166 (0.002)	−0.280 (0.001)	0.246 (0.001)
TYPE			1.000 (0.000)	−0.585 (0.001)	0.451 (0.001)
SIZE				1.000 (0.000)	−0.630 (0.000)
GROWTH					1.000 (0.000)

* Significance levels in parentheses

DILUTE is the complement of the proportion of equity retained by the existing stockholders. *SIZE* is the log of the post-offering size, measured as the sum of the assets at the time of the IPO and the increase in assets from the IPO. This measure is only an approximation of post-offering size, because the assets shown on the IPO prospectus are measured at the most recent fiscal year-end rather than on the offering date. *GROWTH*, or change in size, is measured as the ratio of the increase in the client firm's assets from the offering to the post-offering size. The numerator is computed by adjusting the proceeds for secondary distributions to owners.

Table 4 presents a correlation matrix for the independent variables used in the regression. Table 5 summarizes the results of the logistic regression. Three independent variables are significant: investment banker class, offering type, and post-offering size. The signs of the coefficients are in the anticipated direction in each case. Clients of prestigious investment bankers appear to be more likely to switch to reputable CPAs prior to the IPO than do clients of smaller investment bankers. The displacement of a small auditor with a larger one is also more likely for FC offerings, supporting the position that the investment banker influences the client's choice of auditor in an important way. *DILUTE* is not significant, and in addition carries a negative coefficient, questioning the importance of the monitoring costs argument. Of the two variables representing the demand for control, *SIZE* was significant while *GROWTH* was not. The significance of *SIZE* is consistent with the argument that the demand for control changes at the time of the public offering. However, size may surrogate for factors other than control needs.

We also examine the post-IPO switches for the companies in our sample. If the relative benefits of using a more prestigious auditor diminish following the IPO, then we would expect to find a large number of clients switching to less credible, but less expensive auditors subsequent to the offering. We are able to identify the auditor of record at the end of the calendar year subsequent to the IPO year (i.e., 1986 for 1985 IPOs, 1987 for 1986 IPOs) for 1,014 of the companies in our sample. The auditor information is ob-

Table 5
Logistic Regression Results for Auditor Choice Model
(n±342)

	Coefficient
INTERCEPT	−1.587
IBANK	−0.778**
DILUTE	−0.004
TYPE	−1.432**
SIZE	0.501*
GROWTH	0.290

Model chi-square = 101.28
(with 4 degrees of freedom)**

$R = 0.578$[a]

* Significant at the 0.05 level
** Significant at the 0.01 level
[a] R is the value such that $R^2 = [L(0) - L(M) - p]/L(0)$, where $L(M)$ is the log-likelihood of the model at convergence, $L(0)$ is the log-likelihood with only the intercept in the model, and p is the number of variables in the model.

tained from 10-Ks, 8-Ks, proxies, and some secondary sources like the Compact Disclosure data base. One hundred ninety firms (18.7 percent of the sample) changed auditors in the period examined. Of these changes, only 35 (or 18.4 percent) are to a less credible auditor than the one employed for the IPO. The majority of changes are within the same auditor category (102, or 53.7 percent), while 53 firms (27.9 percent) selected a more credible auditor. These proportions are very close to those cited previously for the population at large. However, the frequency of changes following the IPO is surprisingly high.[16]

Auditor Selection and Investment Banker Fees

In section I, we argue that the investment banker's risk decreases as the credibility of the auditor associated with the IPO increases. This should lead to underwriters charging lower fees for clients audited by the more prestigious auditors, other things being equal. We test this proposition empirically using the following regression model of investment banker compensation:

$$COMP = b_0 + b_1(SDRET) + b_2(PROCEEDS) + b_3(DILUTE) + b_4(AGE) + b_5(IBANK) + b_6(CPATYPE1) + b_7(CPATYPE2) + e,$$

where:

$COMP$ = investment banker compensation;
$SDRET$ = the standard deviation of the firm's returns in the after-market;
$PROCEEDS$ = the log of the proceeds from the offering;

---

[16] We do not investigate the reasons for this high frequency of auditor changes subsequent to the IPO. It appears that the IPO alters the relative demand for different audit product attributes. Prior studies have documented reasons for firms to change auditors, such as financial distress (Schwartz and Menon 1985), agency costs (Francis and Wilson 1988), and supplier cost advantages (Johnson and Lys 1990). It is possible that some of these factors come into play subsequent to the public offering.

DILUTE = the percentage of stock ownership dilution caused by the offering;

AGE = the age of the client;

IBANK = investment banker class, rated from 1 to 4, where 1 is the most prestigious;

*CPATYPE*1, *CPATYPE*2 = dummy variables representing the class of CPA employed by the firm. These variables are assigned values of (0, 0) if the CPA is a Big Eight firm, (1, 0) if the CPA is a national firm, and (0, 1) if the CPA is a local firm.

The dependent variable (*COMP*) is computed as the ratio of investment banker compensation to issue proceeds. Investment banker compensation is measured as the difference between the amount paid by the public for the stock issued and the amount paid by the underwriter to the issuer. Compensation and proceeds data are obtained from the ROS tape.

Like prior studies on underwriter compensation (e.g., Bhagat and Frost 1986; Booth and Smith 1986), we contend that investment banker fees increase with the riskiness of the client. *SDRET* is a market-based measure of risk. Since the securities of the firms in the sample consist of unseasoned common stock, any market-based risk measures can be computed only on an *ex post* basis, using market price data from the period subsequent to the public offering. The measure is flawed not only in that it may reflect factors prevalent at a time subsequent to the public issue but because the market's assessment of the riskiness of the firm subsequent to the issue could be affected by the success of the issue itself. We choose to include *SDRET* because of the importance of incorporating some measure of the market's perceptions of risk. *SDRET* is calculated as the standard deviation of the firm's weekly stock price returns over the 52-week period following the public offering.[17] Weekly stock returns data are obtained from the CRSP NASDAQ file and Standard & Poor's Daily Stock Price Record.

Economies of scale for the underwriter would result in the investment banker compensation decreasing with increases in the size of the offering. In order to control for this potential decrease, the size of the issue (*PROCEEDS*) is included as an independent variable. This variable may also surrogate for riskiness of the issue (Booth and Smith 1986), which should yield the same negative coefficient. *PROCEEDS* is computed as the log of the proceeds from the issue.

The percentage of dilution caused by the IPO (*DILUTE*) is also used as an explanatory variable. If the entrepreneur's shareholdings provide information to investors on the IPO's valuation and thus serve to reduce the information asymmetry between the investor and the issuer (Leland and Pyle 1977), there should be a corresponding reduction of underwriting costs as retained holdings increase. *DILUTE* is obtained from the ROS tape.

---

[17] There is some disagreement among the studies modeling underwriting costs as to which market risk measure should be used. Booth and Smith (1986) use both the systematic component and the unsystematic component (as a proportion of the systematic component) of the total variance of equity returns as independent variables. Bhagat and Frost (1986), who study costs of seasoned equity issues, use the standard deviation of market returns and beta. The first of these measures is used more as a proxy for information costs than as a proxy for risk. However, the authors note that beta may not be an appropriate measure of risk for an underwriter. The principal owners of the investment banking firms have a substantial portion of their wealth in the firm, and are unlikely to be able to diversify away all of the unsystematic risk associated with holding the unpurchased shares from the offering. Johnson and Miller (1988) also used the standard deviation of returns in the after-market.

The information asymmetry problems in an IPO are likely to be less severe for older firms than for those that are relatively new. Older firms generally are better known and have more of a track record in terms of management's credibility.[18] Investment bankers are likely to compensate themselves for the greater risk posed by newer firms by charging a higher fee. The variable *AGE* represents the age of the firm, measured from the date of incorporation. Because of a potential outlier problem, *AGE* is truncated at 20 years. The date of incorporation is obtained from the ROS tape, *Moody's Annual Industrial Manuals*, and 10-K reports.

IBANK is an ordinal variable used to represent the class of investment banker employed by the issuing firm. Prestigious investment bankers are likely to charge a premium for the additional credibility and resources that they bring to a public offering, which should result in a negative relationship between the compensation and *IBANK*. However, the average spread charged by prestigious underwriters for their clients is generally lower than what less prestigious underwriters charge. This anomaly is due to differences in the quality of client pools. Prestigious underwriters are associated with less risky clients (Johnson and Miller 1988). The investment banker variable is likely to serve as a surrogate for the quality of the offering.

The dummy variables for the category of CPA firms (*CPATYPE*1 and *CPATYPE*2) are used to capture the effect of the auditor type on the underwriting costs. These are the variables of interest in this analysis. They are expected to be positively related to *COMP* under the credibility hypothesis.

Separate analyses are performed for FC and BE offerings. As discussed in section I, these two types of contracts carry significantly different types of risk for the investment banker. Both BE and FC contracts contain the risk of reputation loss and of potential litigation on account of offering-related misstatements. However, in FC contracts the underwriter is obliged to retain unsold shares. The extent to which each independent variable affects the investment banker's compensation is likely to be a function of the type of contract employed. For example, while the *ex ante* riskiness of the offering is expected to be reflected in the investment banker's fees in both FC and BE cases, we would expect the magnitude of the effect to be greater for FC contracts.

Of the 800 FC and 520 BE firms identified for the 1985–1986 period, complete data are available only for 611 FC and 109 BE firms. The principal constraint is *SDRET*, which requires complete stock price data for the period following the offering. Many of the companies in our sample are not included in the CRSP file or Standard & Poor's Daily Stock Price Record. The BE sample, in particular, includes a large number of start-ups and other small and somewhat obscure firms. Correlation matrices are presented in table 6.[19] The results of the regressions are presented in table 7.

The model for FC companies has extremely good fit (adjusted $R^2=0.70$), while the model for BE companies performs less well (adjusted $R^2=0.135$). *PROCEEDS* is clearly significant for both groups of companies, reflecting the expected effect of economies of scale. *DILUTE* is also significant in both regressions, consistent with the argument that the percentage of dilution is a signal of the quality of the issue.

---

[18] As an example, our sample of IPOs include Fireman's Fund, a firm with a well-known record.

[19] The high correlations between the independent variables indicate a potential multicollinearity problem. We attempt to assess the severity of the problem by computing variance inflation factors (see, e.g., Kennedy 1985, 153). The magnitude of the variance inflation factors is within acceptable limits, suggesting that harmful collinearity may not be present.

Table 6
Correlation Matrices for Underwriter Compensation Model*

Panel A. Firm Commitment Sample (*n*=611):

	SDRET	PROCEEDS	DILUTE	AGE	IBANK	CPATYPE1	CPATYPE2
SDRET	1.000 (0.000)	−0.292 (0.001)	−0.066 (0.103)	−0.213 (0.001)	0.271 (0.001)	0.013 (0.750)	0.067 (0.097)
PROCEEDS		1.000 (0.000)	0.027 (0.507)	0.221 (0.001)	−0.671 (0.001)	−0.041 (0.316)	−0.262 (0.001)
DILUTE			1.000 (0.000)	−0.174 (0.001)	0.072 (0.074)	−0.002 (0.952)	0.047 (0.249)
AGE				1.000 (0.000)	−0.239 (0.001)	−0.081 (0.841)	−0.053 (0.188)
IBANK					1.000 (0.000)	0.089 (0.028)	0.227 (0.001)
CPATYPE1						1.000 (0.000)	−0.117 (0.004)
CPATYPE2							1.000 (0.000)

Panel B. Best Efforts Sample (*n*=109):

	SDRET	PROCEEDS	DILUTE	AGE	IBANK	CPATYPE1	CPATYPE2
SDRET	1.000 (0.000)	−0.050 (0.605)	−0.238 (0.012)	−0.102 (0.293)	0.078 (0.420)	0.103 (0.286)	0.144 (0.136)
PROCEEDS		1.000 (0.000)	0.211 (0.028)	−0.065 (0.501)	−0.033 (0.731)	0.063 (0.516)	−0.209 (0.030)
DILUTE			1.000 (0.000)	−0.008 (0.935)	0.087 (0.370)	−0.171 (0.076)	0.060 (0.529)
AGE				1.000 (0.000)	−0.046 (0.633)	−0.032 (0.743)	0.004 (0.965)
IBANK					1.000 (0.000)	0.219 (0.022)	−0.057 (0.558)
CPATYPE1						1.000 (0.000)	−0.441 (0.001)
CPATYPE2							1.000 (0.000)

* Significance levels in parentheses

IBANK, the investment banker variable, is positively related to the dependent variable *COMP* for FC companies, reflecting the expected quality differences between the groups of companies underwritten by the different investment bankers. A likely explanation for the lack of significance of this variable in the BE regression is that no major or sub-major investment bankers (categories 1 and 2) are associated with BE offerings. Interestingly, *SDRET*, the market measure of risk, is also significantly related to the dependent variable *COMP* only for FC companies.

Table 7
Regression Results for Underwriter Compensation Model

	Coefficient (t-value)		
	FC Firms (n=611)	*BE Firms* (n=109)	*BE Firms (omitting* SDRET) (*n*=417)
Intercept	12.808 (25.72)**	12.577 (13.10)**	11.108 (16.54)**
SDRET	5.903 (6.02)**	0.069 (0.06)	—
PROCEEDS	−0.711 (15.52)**	−0.376 (3.95)**	−0.265 (5.02)**
DILUTE	0.012 (6.11)**	0.010 (2.63)*	0.008 (3.14)**
AGE	−0.004 (0.81)	−0.003 (0.28)	−0.002 (0.21)
IBANK	0.406 (10.90)**	−0.044 (0.30)	0.105 (0.77)
CPATYPE1	0.211 (2.05)*	0.255 (1.41)	0.178 (0.99)
CPATYPE2	0.259 (2.20)*	0.232 (1.65)	0.322 (2.44)*
R^2 (adjusted)	0.700	0.135	0.125
F-value	204.35**	3.40**	10.88**

* Significant at the 0.05 level
** Significant at the 0.01 level

The variables of interest in this study, the dummy variables for CPA type, are significant in the case of the FC regression. The coefficient is positive in each case, suggesting that clients of these firms are charged a premium underwriting fee over that charged to clients of Big Eight CPA firms. An *F*-test for differences between b_6 and b_7 shows no difference (*F*-value=0.12, p=0.74), indicating the absence of a discount for firms audited by national CPAs, compared to firms audited by a local CPA.

In the case of the BE sample, the CPA dummies are not significant. The most plausible explanation for this finding, consistent with the lack of significance for *SDRET*, is that the investment banker's economic risk is less for the BE sample. Not only is the underwriting risk absent, but the risk of potential reputation loss that is normally associated with all offerings is likely to be less evident in our BE sample, since there are no offerings underwritten by major or sub-major underwriters.[20]

---

[20] Instead of running separate regressions for FC and BE firms, it is possible to use a single regression model that includes one dummy variable for offering type and interaction terms to capture the change in the slope for each independent variable between FC and BE regimes. Such a regression would assume that the variance of the error terms would be the same for FC and BE cases, which does not hold for our sample. None-

The BE regression is limited by the reduction in sample size imposed by *SDRET* data requirements. In order to test the sensitivity of the results to this loss of observations, we rerun the regression, dropping *SDRET*. The sample size increases from 109 to 417. Table 7 reports the results of this regression. This time *CPATYPE2*, the dummy variable representing the use of a local CPA firm, is significant. The results of this regression need to be interpreted keeping in mind that *SDRET* has been omitted and that *CPATYPE2* may serve as a surrogate for risk. *F*-tests show no differences between the coefficients for the two CPA dummies for either of the two best efforts regressions (sample of 109 firms: *F*-value=1.62, p=0.20; sample of 417 firms: *F*-value=0.75, p=0.39).

We view the overall results as supporting the auditor credibility hypothesis. Investment bankers appear to adjust their fees for FC offerings to compensate for the level of credibility of the client's auditor. There is no evidence of such adjustments in the case of BE, where the investment banker's incentives to require more credible auditors are weaker as well. Again, this suggests that the insurance demand for credible auditors may be weak.

III. Summary and Conclusions

This study investigates the hypothesized preference of investment bankers and their clients for reputable auditors at the time of the initial public offering. We find fewer instances than expected of auditor switching in companies about to go public. However, most of the changes that take place lead to small auditors being displaced by larger, more reputable ones, consistent with the auditor credibility hypothesis. A logistic regression analysis provides additional evidence that credibility is the reason for the replacement of small auditors. Clients that select larger auditors tend to be those that employ prestigious investment bankers and make "firm commitment" offerings. They also tend to be larger than those firms that continue to use small auditors, suggesting the presence of incentives to switch auditors for reasons other than credibility.

A cross-sectional analysis of the fees charged by the investment bankers provides strong support for the auditor credibility hypothesis. Investment banker compensation is found to be affected by the firm's auditor in the case of "firm commitment" offerings, where there appears to be a "discount" for Big Eight audited firms. Surprisingly, we are unable to detect any differences between compensation charged to clients of national CPA firms and clients of local firms. The evidence of an auditor effect on compensation is weaker for "best efforts" offerings, where the investment banker's incentives to prefer a more credible auditor are also weaker.

This paper adds to the body of evidence that indicates that the supply of auditing services is not homogeneous. Despite uniform certification requirements and generally accepted auditing standards, it appears that users of audited financial statements perceive auditors as having different levels of credibility and make economic decisions on

---

theless, to satisfy ourselves that the type of contract makes a difference, we regressed *COMP* on the independent variables listed in table 7, adding a dummy variable for offering type (*OFFTYPE*) and the interaction terms *OFFTYPE*SDRET*, *OFFTYPE*PROCEEDS*, *OFFTYPE*DILUTE*, and *OFFTYPE*AGE*. The dummy variable was not significant. However, two of the interaction terms (*OFFTYPE*SDRET* and *OFFTYPE*PROCEEDS*) were significant, indicating that magnitude of the effect of *SDRET* and *PROCEEDS* on *COMP* depends on the offering arrangement.

this basis. It is likely that the importance of auditor credibility is contingent upon the credibility of the client's management and the availability of other sources of information about the firm. Firms making IPOs tend to be small and relatively obscure, thus providing a context in which the auditor's certification of the accuracy of management's disclosures is particularly important. We believe that it is important to identify other contexts in which these conditions of low management credibility and lack of alternative sources of information exist, and to empirically examine the effect of auditor choice in these contexts. It is also important to theoretically model more completely the conditions that make auditor credibility important.

References

Bachar, J. 1989. Auditing quality, signaling, and underwriting contracts. *Contemporary Accounting Research* 6 (Fall): 216–41.

Balvers, R., B. McDonald, and R. E. Miller. 1988. Underpricing of new issues and the choice of auditor as a signal of investment banker reputation. *The Accounting Review* 63 (October): 605–22.

Beatty, R. 1989. Auditor reputation and the pricing of initial public offerings. *The Accounting Review* 64 (October): 693–709.

——, and J. Ritter. 1986. Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics* 15 (January/February): 213–32.

Berton, L. 1986. Hunting for small game. *The Wall Street Journal* 67 (May 19): 25D.

Bhagat, S., and P. Frost. 1986. Issuing costs to existing shareholders in competitive and negotiated underwritten public utility equity offerings. *Journal of Financial Economics* 15 (January/February): 233–60.

Booth, J., and R. Smith. 1986. Capital raising, underwriting and the certification hypothesis. *Journal of Financial Economics* 15 (January/February): 261–81.

Carpenter, C., and R. Strawser. 1971. Displacement of auditors when clients go public. *Journal of Accountancy* 131 (June): 55–58.

DeAngelo, L. 1981a. Auditor independence, "low balling," and disclosure regulation. *Journal of Accounting & Economics* 3 (August): 113–27.

——. 1981b. Auditor size and audit quality. *Journal of Accounting & Economics* 3 (December): 183–99.

Francis, J., and D. Simon. 1987. A test of audit pricing in the small-client segment of the U.S. audit market. *The Accounting Review* 62 (January): 145–57.

——, and E. Wilson. 1988. Auditor changes: A joint test of theories relating to agency costs and auditor differentiation. *The Accounting Review* 63 (October): 663–82.

Hall, W., and A. Renner. 1988. Lessons that auditors ignore at their own risk. *Journal of Accountancy* 166 (July): 50–58.

Jacobs, S. 1983. Small CPA concern sues an underwriter over loss of client. *The Wall Street Journal* 64 (July 18): 4.

Johnson, J., and R. Miller. 1988. Investment banker prestige and the underpricing of initial public offerings. *Financial Management* 17 (Summer): 19–29.

Johnson, W. B., and T. Lys. 1990. The market for audit services. *Journal of Accounting & Economics* 12 (January): 281–308.

Kennedy, P. 1985. *A Guide to Econometrics*. Cambridge, MA: The MIT Press.

Leland, H., and D. Pyle. 1977. Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance* 32 (May): 371–87.

Mandelker, G., and A. Raviv. 1977. Investment banking: An economic analysis of optimal underwriting contracts. *The Journal of Finance* 32 (June): 683–94.

Palmrose, Z. 1986. Audit fees and audit size. *Journal of Accounting Research* 24 (Spring): 97–110.

Public Accounting Report 8. 1985. (December): 5.

Public Accounting Report 9. 1986. (January): 3.

Public Accounting Report 10. 1987. (February 1): 1.

Rock, K. 1986. Why new issues are underpriced. *Journal of Financial Economics* 15 (January/February): 187–212.

Schwartz, K., and K. Menon. 1985. Auditor switches by failing firms. *The Accounting Review* 60 (April): 248–61.

Simunic, D., and M. Stein. 1987. *Product Differentiation in Auditing: Auditor Choice in the Market for Unseasoned New Issues.* Vancouver, B.C.: The Canadian Certified General Accountants' Research Foundation.

Sutton, D., and M. W. Benedetto. 1988. *Initial Public Offerings: A Strategic Planner for Raising Equity Capital.* Chicago, IL: Probus Publishing.

Tinic, S. 1988. Anatomy of initial public offerings of common stock. *The Journal of Finance* 43 (September): 789–822.

Titman, S., and B. Trueman. 1986. Information quality and the valuation of new issues. *Journal of Accounting & Economics* 8 (June): 159–72.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 333–346

Effectiveness of Rectification in Audit Sampling

John Wurst
Emory University
John Neter
University of Georgia
James Godfrey
George Mason University

SYNOPSIS: This article considers audit sampling and the auditor's risks of incorrect conclusions in deciding whether an accounting population contains a material error amount. The proposed new sample evaluation procedure takes into account the correction of all errors observed in the sample and calculates a modified upper bound on the total error amount remaining in the unsampled population. This procedure is known as rectification. It is shown that rectification significantly reduces the risk of incorrect rejection when sample sizes of 150 and 300 are used.

Since the size of the error amount observed in the sample is crucial to the extent of risk reduction, the first stage of this study investigated the magnitude of rectification for three dollar-unit sample selection methods commonly used in practice—systematic, random, and cell selection. Theoretical results for the expected error amount contained in systematic samples were obtained for 32 different study populations for three sample sizes (65, 150, and 300).

Simulation was used to estimate the expected error amounts with random and cell selection for the same 32 study populations and three sample sizes. The results show that the average error amount subject to rectification with cell selection is very close to that with systematic selection for all three sample sizes. Random selection produced smaller expected error amounts than did systematic selection, especially for the larger sample sizes.

In the second part of the study, the risks of incorrect rejection with and without rectification were compared. Factors considered in this com-

We are greatly indebted to the referees for most helpful comments that led to major extensions of the analysis and an improved framework of presentation.

Submitted August 1989.
Revisions received May, July, and September 1990.
Accepted October 1990.

parison were sample size (65, 150, or 300) and sampling method (systematic, random, or cell).

Use of rectification led to noticeable reductions in the risk of incorrect rejection for sample sizes other than 65. For a given application, auditors will need to determine whether the cost of the additional sampling is balanced by the savings from the reductions in the risk of incorrect rejection with rectification.

Key Words: *Dollar-unit sampling, Rectification, Sampling risks, Stringer bound.*

Data Availability: *More detailed tables containing additional results are available from the authors.*

IN performing substantive tests for accounting populations, auditors often employ statistical sampling methods to make inferences about the total dollar amount of errors in the population. None of these methods, however, recognizes the potential for using the information about errors identified in the sample to make inferences only about the dollar amount of errors remaining in the population. Royall (1970) suggested this inference approach for estimation for finite populations as more precise because the only uncertainty is for the unsampled population elements since the characteristics of the sampled elements are known with certainty. This suggestion is particularly applicable to dollar-unit sampling in auditing where the audit provides information not only about the errors in each sampled dollar but also about the errors in the other dollars of the line items brought into the sample.

Correcting all errors found in the sampled line items and making inferences about the dollar amount of errors remaining in the population is called rectification. The purpose of rectification is to reduce the risks of incorrect decisions when the auditor must decide, on the basis of sample evidence, whether a material amount of errors is present in the population. The potential use of rectification in auditing has been recognized by Leslie and Aldersley (1982) and by Garstka (1982). Indeed, Garstka suggested that the impact of rectification on the risks of incorrect decisions could be substantial since rectification of errors could make a population with a materially large amount of errors acceptable after rectification.

In this article, we describe an investigation of the effectiveness of rectification in reducing the risks of incorrect rejection, since auditors typically control the risk of incorrect acceptance of a population with a material amount of errors. In section I, we discuss rectification in more detail. Section II presents theoretical results on the magnitude of error that can be identified in a sample. This section also presents data on the actual magnitudes of the total error amounts identified in samples for a variety of study populations and sample sizes for three sampling procedures commonly used in dollar-unit sampling. In section III, we present results on the reduction of the risks of incorrect rejection with rectification for a variety of study populations and sample sizes. We conclude with a brief summary and discussion in section IV.

I. Rectification in Auditing

Purpose of Audit Rectification

In current practice, when auditors use statistical sampling to examine an accounting population for purposes of determining whether it contains a material dollar amount of errors, the errors found in the sample are used basically for obtaining an upper bound on the total amount of errors in the population. To our knowledge, corrections of nonmaterial error amounts are generally left to the discretion of the client. This practice creates no problems for the auditor in developing a conclusion about the total account balance under test since the statistical inference pertains to the total dollar amount of errors in the population, including the observed nonmaterial amounts in the sample. Material errors found in the sample do, of course, require corrections.

With rectification, conversely, all errors observed in the statistical sample would need to be corrected in the individual accounts because, with rectification, the auditor's objective is to make inferences about the total amount of errors remaining in the unsampled portion of the population. As the target of inference is reduced in this fashion, the risks of incorrect decisions can be reduced when auditors make judgments about whether the population contains a material amount of errors.

Because the emphasis with rectification is on making inferences about the unsampled portion of the population, the inference procedures currently used by auditors require some modifications, but no additional information is needed. First, the errors observed in the sample are corrected. Second, a modified upper confidence bound for the total dollar amount of errors remaining in the unsampled population is constructed by subtracting the error amount observed in the sample from the conventional upper bound for the total error amount in the population (sampled and unsampled items). Finally, the modified bound is compared in the usual fashion with the amount deemed to be material for the account being tested. If the modified bound equals or exceeds materiality, then more work would be required. Otherwise, the account balance would be acceptable.

Relation of Audit Rectification to Booking of Adjustments

Rectification must be distinguished from the booking of an adjustment to a total account balance. Booking of adjustments occurs after the auditor has decided, on the basis of a statistical sample or otherwise and after additional audit work, that a material amount of errors is present and that the book amount needs to be adjusted to represent fairly the financial position and the results of operations. This decision typically is made relative to the aggregation of all of the main accounts that make up the financial statements, including those that are audited without statistical sampling.

The *AICPA Professional Standards* discusses this situation in the context of the auditor's overall "... accumulation of sufficient evidential matter." If, in this accumulation, "... the aggregation of likely errors causes the financial statements to be materially misstated, he should request management to eliminate the material misstatement." An upper bound on the total error amount calculated from sample results for a statistically sampled account is defined to be part of the "likely errors," and "material misstatements may be eliminated by ... application of appropriate accounting principles, other adjustments in amounts, or the addition of appropriate disclosure of inadequately disclosed matters" (AICPA 1988, AU § 312.31).

Both the AICPA standard and extant literature (e.g., Arens and Loebbecke 1988; Guy and Carmichael 1986) imply that adjustments at an aggregate level are acceptable. Although the required correction of all errors observed in a sample with rectification differs from current audit practice, it is not in conflict with any aggregate adjustment mandated by the AICPA.

II. Extent of Rectification with Sampling

Since the magnitude of the reduction in the risks of incorrect decisions is directly influenced by the dollar amount of errors that are observed in the sample, the first part of this investigation concerns the dollar amount of errors found in the sample when dollar-unit sampling is employed. We first present some theoretical results for systematic sampling and then provide data on actual magnitudes for various study populations and sample sizes. We conclude this section by presenting comparable results for random and cell sampling of dollar units.

To simplify the exposition, we assume that all errors are overstatement errors. Overstatement errors are often viewed by auditors as more serious than those of understatement, and they frequently are the dominant type of error (e.g., Johnson et al. 1981). Furthermore, the offsetting of over- and understatement errors is not meaningful here since we are interested in the total amount of rectification that can take place on the basis of the errors observed in the sample. Thus, it would be preferable when understatement errors are present to consider the total absolute error amount. The results to be presented can be viewed as representing the total absolute error amount when understatement errors are present.

Theoretical Results for Simple Random Sampling of Line Items

We first consider the extent of rectification with simple random sampling of line items without replacement. We shall denote the number of line items in the population by N, the book (recorded) amounts of the line items by $Y_1, Y_2, \ldots, Y_N$, and the total book amount by $T(Y)$:

$$T(Y)=\sum_{i=1}^{N} Y_i. \tag{1}$$

The error amounts of the N line items in the population will be denoted by $E_1, E_2, \ldots, E_N$, the total error amount by $T(E)$.

$$T(E)=\sum_{i=1}^{N} E_i. \tag{2}$$

The mean error amount per line item in the population is denoted by $\bar{E}$.

With a simple random sample of n line items selected without replacement, the expected error amount in the sample subject to rectification (by definition of expected value) is as follows:

$$\text{Expected error amount}=\sum_{i=1}^{N} E_i\left(\frac{n}{N}\right)=\frac{n}{N}T(E)=n\bar{E}. \tag{3}$$

If simple random sampling of line items is done with replacement, the expected error amount in the sample will be smaller than the amount in equation (3) because a given line item may enter the sample more than once.

From equation (3) it is clear that as the sample size n and/or the mean error amount $\bar{E}$ become larger, the larger will be the expected error amount in the sample to be rectified. Further, equation (3) implies that the distribution of errors among the line items does not affect the expected error amount in the sample.

Theoretical Results for Systematic Dollar-Unit Sampling

In contrast to random sampling of line items, the expected error amount with systematic dollar-unit sampling (DUS) depends on how the errors are distributed among the line items. This is because systematic DUS is a probability-proportional-to-size (pps) sampling procedure, the size variable being the book value.

We shall assume throughout that no line item in the population is larger than the sampling interval for systematic sampling; that is, we shall assume that $Y_i \le T(Y)/n$, which is consistent with auditors' examining all large line items. When no line item in the population exceeds the sampling interval, systematic DUS selects line items without replacement, and the probability that the ith population line item will be included in a systematic sample of size n is $nY_i/T(Y)$.[1]

For systematic sampling of dollar units, the expected error amount in the sample subject to rectification is:

$$\text{Expected error amount} = \sum_{i=1}^{N} E_i \left[\frac{nY_i}{T(Y)}\right] = \frac{n}{T(Y)} \sum_{i=1}^{N} E_i Y_i. \tag{4}$$

Equation (4) shows that, for given book and error amounts, the larger the systematic sample size n, the larger will be the expected error amount in the sample subject to rectification. Similarly, the larger the error amounts for a given sample size, the greater will be the expected error amount in the sample.

Rectification with systematic dollar-unit sampling is advantageous when overstatement errors tend to be concentrated in large line items. In that case, systematic dollar-unit sampling will tend to bring in a larger error amount than will simple random sampling of line items. This can be seen by using the definition of the coefficient of correlation between the error and book amounts,

$$\rho_{EY} = \frac{E\{EY\} - \bar{E}\bar{Y}}{\sigma_E \sigma_Y}, \tag{5}$$

where σ_E and σ_Y are the population standard deviations of the two variables E and Y, and $\bar{E}$ and $\bar{Y}$ are the respective population means.

Using equation (5) in equation (4), we obtain:

$$\text{Expected error amount} = \frac{n}{\bar{Y}}(\rho_{EY}\sigma_E\sigma_Y + \bar{E}\bar{Y}). \tag{6}$$

---

[1] Cochran (1977, 261–66) describes several other pps sampling procedures without replacement that have these inclusion probabilities.

Table 1
Book Value Characteristics of the Basic Populations 1M and 4

	Population	
Characteristic	*1M*	*4*
Mean	$40.35	$1,860.39
Standard deviation	$72.32	$3,865.13
Skewness	6.11	3.2
Kurtosis	48.29	11.4
Number of items	8,282	4,033
Total book value	$334,207.70	$7,502,957.20
Maximum	$948.28	$24,928.60
Minimum	$0.50	$0.10

Thus, for given means and standard deviations, the larger the correlation coefficient ρ_{EY}, the greater will be the expected error amount in the sample to be rectified.

The expected error amount in the sample subject to rectification as a percentage of the original total error amount in the population can be obtained from equation (6),

$$\frac{\text{Expected error amount}}{T(E)}(100)=\frac{n}{N}(\rho_{EY}CV_E CV_Y+1)(100). \tag{7}$$

where $CV_E=\sigma_E/\bar{E}$ is the population coefficient of variation of the error amounts E_i and $CV_Y=\sigma_Y/\bar{Y}$ is the population coefficient of variation of the book amounts Y_i.

Extent of Rectification with Systematic DUS

To provide magnitudes of the extent of rectification that may be expected for different types of accounting populations, we created 32 study populations while controlling a number of factors that may affect the extent of rectification. The relevant parameters of the book and error amounts were calculated for each, and the expected percentage of the total error amount subject to rectification with systematic sampling of dollar units was then calculated by means of equation (7) for three sample sizes (65, 150, and 300) in each population. The three sample sizes range from moderate to large.

Creation of Study Populations. The study populations used here are the same ones employed for the Wurst et al. (1989) study. They were created by seeding errors into two actual accounting populations (labeled 1M and 4) from the Neter and Loebbecke (1975) study while controlling for three factors that may affect the extent of rectification—total error amount, bunching pattern, and mean taint setting. We provide a brief description of the study populations now. A complete description of the study populations can be found in Wurst (1987).

Population 1M is an accounts receivable population from a freight company and is characterized by a large number of small accounts. Population 4 is an accounts receivable population obtained from a manufacturing firm, and it consists of a smaller number of large accounts with much more variable book values than those in population 1M. These populations were chosen to provide contrasting characteristics, as summarized in table 1. Neither population contains any line items exceeding any of the sampling intervals considered.

Table 2

Mean Expected Percentage of Total Error Amount Subject to Rectification with Systematic DUS, by Study Population Characteristics

Population Characteristic	*Sample Size*			*Ratio to Simple Random Sampling of Line Items*
	65	*150*	*300*	
Overall	7.1	16.3	32.6	5.74
Bunching pattern				
Uniform	5.6	12.9	25.8	4.52
Increasing	8.5	19.7	39.3	6.96
Accounting population				
1M	4.2	9.6	19.3	5.32
4	9.9	22.9	45.9	6.17
Accounting population —Bunching pattern				
1M-Uniform	3.2	7.4	14.9	4.10
4-Increasing	11.9	27.5	55.0	7.40

Four total error amounts were utilized, representing 1.25, 2.5, 5, and 10 percent of the population total book value.

Bunching pattern refers to the correlation between the error amount and the book value of line items. Two bunching patterns were employed, an increasing pattern and a uniform pattern. The increasing pattern concentrated 80 percent of the total error amount in the group of large line items, with the remaining 20 percent in the group of small line items. The uniform pattern divided the total error amount equally between the two groups of line items. The two groups of line items were formed by sorting the line items by book value and dividing the population into two groups of approximately equal total book value.

The two groups of line items used for controlling the bunching pattern also were employed for controlling the mean taint setting. Two levels of mean taints were used, low and high. These mean taint settings were controlled as follows:

Mean Taint Level	*Population Group*	
	Small Book Value Line Items	*Large Book Value Line Items*
Low (truncated exponential distribution)	0.35	0.25
High (uniform distribution)	0.70	0.50

The exponential distribution for the low mean taint level was truncated at 1.0. The taint distributions for the small line items also include a discrete probability mass at 1.0. Seeding taints in this fashion was designed to produce study populations with characteristics consistent with current evidence about accounting populations (e.g., Johnson et al. 1981).

Findings. Table 2 contains means of the percentages of expected error amounts subject to rectification by systematic sampling of dollar units for a given sample size.

Thus, each entry in table 2 is an average percentage over the relevant study populations. As shown, the overall average extent of rectification is 7.1 percent for sample size 65 and increases to 32.6 percent for sample size 300.

As expected from equation (7), the percentage of the total error amount that can be rectified with systematic DUS increases proportionally with the increase in sample size. The percentage of the total error amount that can be rectified is neither affected materially by the error amount in the population nor by the mean taint level. We, therefore, do not report the detailed results.

Table 2 shows that the effect of the bunching pattern is consistent with theoretical expectations. The increasing bunching pattern led to a larger correlation coefficient ρ_{EY} and also to a larger coefficient of variation CV_E. Hence, equation (7) indicates that, for given mean taint and total error amount, the percentage of the total error amount that is subject to rectification should be higher for the increasing than for the uniform bunching pattern. The results in table 2 are entirely consistent with this expectation. This difference becomes larger with increasing sample size.

The effects of accounting population also are consistent with expectations. Since the coefficients of variation CV_Y and CV_E are quite similar for the two accounting populations, as are the coefficients of correlation ρ_{EY}, the percentage of the total error amount subject to rectification will be much greater according to equation (7) for accounting population 4 than for accounting population 1M (for given sample size) because population 1M has twice as many line items as population 4. Table 2 bears this out.

The main effects for bunching pattern and accounting population shown in table 2 do not fully disclose the magnitudes of their combined effects because these factors interact in their effects on the rectification percentage. Table 2 provides information on this interaction by presenting results for two extreme cases—accounting population 1M with a uniform bunching pattern and accounting population 4 with an increasing bunching pattern. The percentage of the total error amount subject to rectification with systematic DUS for population 4 with an increasing bunching pattern is three to four times as large as that for population 1M with a uniform bunching pattern. Thus, even with sample size 65, auditors can identify 11.9 percent of the total error amount when sampling a population that has overstatement errors concentrated in the larger line items, like population 4. This percentage increases to 27.5 percent for sample size 150, and to 55 percent for sample size 300.

The effectiveness of systematic DUS compared with simple random sampling of line items may also be seen in the last column of table 2. The ratios of the expected percentages subject to rectification for the two methods of sampling are constant for all sample sizes, as may be seen by comparing equations (3) and (6). Overall, for each sample size, systematic DUS revealed 5.74 times as many dollar errors as did simple random sampling of line items for the study populations under consideration. This ratio became as high as 7.4 when accounting population 4 with the increasing bunching pattern was sampled.

Extent of Rectification with Random and Cell DUS

Two other methods of pps sampling widely used in auditing are random and cell sampling of dollar units (Leslie et al. 1979). With cell sampling, the population of $T(Y)$ dollar units is divided into n cells of equal sizes, from each of which one dollar unit is

Table 3
Average Total Error Amount Subject to Rectification with Random and Cell DUS as Percentage of Expected Total Error Amount with Systematic DUS (Averages are over 32 study populations)

Sample Size	*Random Sampling*	*Cell Sampling*
65	93.9	100.1
150	87.2	98.1
300	77.0	94.0

selected randomly and independently. Unlike systematic DUS when no line item exceeds the sampling interval, random sampling of dollar units without replacement and cell sampling of dollar units are, in effect, pps sampling-with-replacement methods because both permit the same line item to be sampled more than once. This can occur when two different selected dollar units pertain to the same line item. Cell sampling is a pps sampling-without-replacement method in one special case, namely, when no line item straddles two cells.

Consequently, the theoretical results obtained earlier can only serve as upper bounds on the expected error amount subject to rectification with random and cell dollar-unit sampling. Instead, simulation has been utilized to obtain information about the expected error amounts subject to rectification with random and cell DUS sampling. Samples of sizes 65, 150, and 300 were selected from each of the 32 study populations and simple random and cell sampling of dollar units were used. Simple random selection of dollar units was carried out with replacement to simplify the computer simulations. However, this procedure is almost the equivalent of selecting the dollar units without replacement because of the large number of dollar units in the study populations relative to the sample sizes.

Five hundred replications were made for each study population, sampling method, and sample size. For each replication, the total error amount observed in the sample was then obtained.

Table 3 summarizes the results of this simulation study. For each sample size, average results of the mean error amount subject to rectification with random and cell DUS are shown as a percentage of the expected error amount obtainable with systematic sampling. For example, the entry 93.9 for random DUS with $n=65$ indicates that the mean error amounts in random dollar-unit samples of size 65 from these study populations were, on average, 93.9 percent as great as the expected amounts for systematic sampling. Thus, random sampling of size 65 involves here only a small amount of sampling with replacement.

The results also show that cell sampling is almost as effective for all sample sizes as systematic sampling in sweeping errors into the sample. Thus, cell sampling involves relatively little sampling with replacement, even for sample size 300.

In contrast, random sampling is less effective than either systematic or cell sampling, especially for sample sizes 150 and 300. For sample size 300, the mean error amount subject to rectification with random sampling is only 77 percent as great as that with systematic sampling.

Table 3 also shows that random and cell sampling become relatively less effective than systematic sampling in bringing errors into the sample as the sample size increases. This is to be expected since the probability of any line item's being included more than once under with-replacement sampling methods increases with larger sample sizes. The effect is small for cell sampling, but is pronounced for random sampling.

III. Effects of Rectification on Risks of Incorrect Rejection

Auditors conventionally formulate their decision rule in terms of an upper bound for the total amount of errors in the population. If this upper bound equals or exceeds the materiality amount, the auditor will conclude that a material error may be present. With this decision-making approach, the risk of incorrect acceptance of a population containing a material error amount is controlled, but not so for the risks of incorrect rejection of populations with error amounts less than the materiality amount. It is, therefore, of interest to examine whether these uncontrolled risks of incorrect rejection are substantially reduced by rectification.

The use of rectification for making inferences about the remaining total amount of errors in the unsampled population requires, as mentioned earlier, that both the conventional upper bound and the decision rule be modified.

Modification of Upper Bound

As stated earlier, we modify a conventional upper bound for the total error amount before rectification, U_b, by subtracting the total error amount observed in the sample, $T(E_s)$:

$$U_a = U_b - T(E_s). \tag{8}$$

If U_b is a valid confidence bound for the total error amount before rectification, $T(E)$, then U_a is a valid confidence bound for the total error amount after rectification, $T(E) - T(E_s)$.[3]

Modification of Auditor's Decision Rule

When rectification is not considered, auditors generally utilize the following decision rule for deciding whether the total error amount $T(E)$ in the population equals or exceeds the materiality level, M:

$$\begin{aligned} &\text{If } U_b \geq M, \text{ conclude that } T(E) \geq M.\\ &\text{If } U_b < M, \text{ conclude that } T(E) < M. \end{aligned} \tag{9}$$

With rectification, the decision rule is as follows:

$$\begin{aligned} &\text{If } U_a \geq M, \text{ conclude that } [T(E) - T(E_s)] \geq M.\\ &\text{If } U_a < M, \text{ conclude that } [T(E) - T(E_s)] < M. \end{aligned} \tag{10}$$

---

[3] Another approach to modifying a conventional bound was used by Smith (1979) and Phillips (1985). They obtained a modified bound for the unsampled population by multiplying the conventional bound for the total population by the proportion of the book amount of the unsampled population to that of the total population. This approach may not yield a valid bound for the total remaining error amount after rectification.

Simulation Study

For the modified decision rule, we need not be concerned with the risk of incorrect acceptance of a population with a material error amount, for two reasons. First, decision rule (10) controls this risk. Second, the coverage properties of the modified bound U_a in equation (8) for the total remaining error amount (i.e., the proportion of times in repeated sampling that the bound U_a will equal or exceed $T(E)-T(E_r)$) are identical to the coverage properties of the conventional bound U_b for the total error amount before rectification. Previous studies have shown that both the Stringer and the cell bounds (Leslie et al. 1979) have good coverage properties with random, cell, and systematic DUS sampling, and perform well with respect to rejecting a materially misstated population (see, e.g., Duke et al. 1985; Plante et al. 1985). Consequently, the modified Stringer and cell bounds with decision rule (10) will bear small risks of accepting a population with a remaining material overstatement error amount.

The uncontrolled risks of incorrect rejection of acceptable populations were investigated in this study for both the Stringer and cell bounds. We report here only the results for the Stringer bound because of its wide use.[3] In particular, we compare the risks of incorrect rejection for the unmodified and modified Stringer bounds when the total error amount in the population before rectification equals 0.5M, for the study populations described previously.

Two materiality levels were selected: $M=2.5$ percent of total book value, and $M=5$ percent of total book value. The eight study populations with total error amount $T(E)=0.0125T(Y)$ correspond to populations with $T(E)=0.5M$ for the 2.5 percent materiality level. Similarly, the eight study populations with total error amount $T(E)=0.025T(Y)$ correspond to populations with $T(E)=0.5M$ for the 5 percent materiality level. Thus, 16 study populations were used in this part of the study.

Two nominal confidence levels were employed in constructing each bound—85 and 95 percent. Only the results for the nominal 95 percent bounds will be presented here because the results for the nominal 85 percent bounds are entirely consistent with the findings for the 95 percent bounds. The sample sizes considered are the same as for the study on the extent of rectification—65, 150, and 300.

For these parameters of the simulation study, the simulations described earlier for the extent of rectification with random and cell sampling were used here in their entirety and were supplemented with comparable simulations for systematic sampling. The risks of incorrect rejection were estimated by the proportion of the 500 replications that led to rejection.

Findings

Table 4 presents aggregate results on the reductions in the risk of incorrect rejection with rectification for the three DUS procedures. The entries are average estimated risks of incorrect rejection for the unmodified and modified Stringer bounds based on a nominal 95 percent level, where the averages are over eight study populations each. In all cases, the error amount in the population before rectification is $T(E)=0.5M$. Also

---

[3] The risks of incorrect rejection for the modified cell bound were lower in this study than those for the modified Stringer bound. This is consistent with the results in Plante et al. (1985) for the unmodified cell and Stringer bounds.

Table 4
Average Risks of Incorrect Rejection for Stringer Bound with Random, Cell, and Systematic DUS for Populations when $T(E)=0.5M$
(Averages are over eight study populations; bounds for nominal 95 percent level)

	$M=0.025T(Y)$			$M=0.05T(Y)$		
Sampling Method	*Unmodified Bound*	*Modified Bound*	*Percent Reduction*	*Unmodified Bound*	*Modified Bound*	*Percent Reduction*
			$(n=65)$			
Random	1.000	1.000	0	0.963	0.961	0.2
Cell	1.000	1.000	0	0.963	0.962	0.1
Systematic	1.000	1.000	0	0.958	0.955	0.3
			$(n=150)$			
Random	0.933	0.927	0.6	0.676	0.586	13.3
Cell	0.946	0.938	0.8	0.691	0.565	18.2
Systematic	0.928	0.922	0.6	0.713	0.586	17.8
			$(n=300)$			
Random	0.673	0.475	29.4	0.253	0.085	66.4
Cell	0.720	0.434	39.7	0.254	0.061	76.0
Systematic	0.727	0.412	43.3	0.222	0.048	78.4

shown in table 4 are the percent reductions in the risk of incorrect rejection with rectification.

There are a number of interesting findings in table 4:

1. For sample size 65, rectification of errors and use of the modified bound achieved little, if any, reduction in the risk of incorrect rejection for any of the three sampling procedures, for either of the two materiality levels.
2. For each sampling procedure, the percent reduction in the risk of incorrect rejection increased with the increase in sample size.
3. The percent reductions in the risk of incorrect rejection tended to be larger for cell and systematic DUS sampling than for random sampling when the sample size is large.
4. The percent reductions in the risk of incorrect rejection with rectification for sample sizes 150 and 300 were consistently larger for the five percent materiality level than for the 2.5 percent materiality level.

None of the population factors had any noticeable effect on the percent reduction of the risk of incorrect rejection for sample size 65, for any of the sampling procedures. For the larger sample sizes, the percent reductions in risk were greater for each sampling method when the errors were bunched in the larger line items, especially when materiality was five percent of the book amount. Also for the larger sample sizes, the percent reductions in risk tended to be much greater for population 4 than for population 1M for all three sampling procedures.

Table 5

Average Risks of Incorrect Rejection for Stringer Bound with Systematic DUS for Populations when $T(E)=0.5M$, by Accounting Population and Bunching Pattern (Averages are over two study populations; bounds for nominal 95 percent level)

	$M=0.025T(Y)$			$M=0.05T(Y)$		
Population-Bunching	*Unmodified Bound*	*Modified Bound*	*Percent Reduction*	*Unmodified Bound*	*Modified Bound*	*Percent Reduction*
			$(n=65)$			
1M-uniform	1.000	1.000	0	0.969	0.969	0
4-increasing	1.000	1.000	0	0.958	0.958	0
			$(n=150)$			
1M-uniform	0.928	0.921	0.8	0.700	0.654	6.6
4-increasing	0.949	0.949	0	0.735	0.427	41.9
			$(n=300)$			
1M-uniform	0.735	0.635	13.6	0.276	0.125	54.7
4-increasing	0.722	0.184	74.5	0.151	0	100.0

We noted in table 2 that bunching pattern and accounting population interact in their effects on the extent of rectification. Table 5 examines this interaction effect on the risk of incorrect rejection for systematic DUS. The effects for the other two sampling methods were similar. We see in table 5 that for the larger sample sizes, especially for materiality at five percent of book value, rectification achieved much greater reductions in the risk of incorrect rejection for population 4 with increasing bunching pattern than for population 1M with uniform bunching pattern. Thus, rectification appears to be most effective in reducing the risk of incorrect rejection when the population contains larger items, the overstatement errors are concentrated in the larger items, and the sample size is not small.

IV. Summary and Discussion

The purpose of this study was to investigate the impact of rectification on reducing the risk of incorrect rejection of an acceptable population. For populations with a total error amount equal to one-half of materiality and sample size 65, no significant reductions in the risk of incorrect rejection were found with rectification. However, for sample sizes 150 and 300, some substantial reductions were observed. For these latter sample sizes, the most pronounced effects occurred when the errors were concentrated in the larger line items, the line items tended to be larger, and the total error amount was not small.

Since a sample size of 65 may be closest to the sizes used in practice, an initial reaction to the findings on risk reduction presented here could be that rectification is not of practical use. However, this reaction would ignore the possibility of a cost/benefit trade-off. Rectification with a sample size of 300 produces a substantial reduction in the risk of incorrect rejection, compared with use of a sample of 65 without rectification. Although there are additional auditing costs associated with a larger sample size, there are also lower expected auditing costs associated with the smaller risks of incorrect rejection. Thus, auditors will need to weigh these two costs. It could be that the reductions in the risks of incorrect rejection may make an enlargement of the sample size worthwhile.

This study should be extended in several directions in future research. First, it would be desirable to make investigations of rectification with additional accounting populations, including ones containing understatement errors. Second, the impact of rectification on the risks of incorrect rejection should be evaluated for other error amounts below the materiality level. This is particularly important for error amounts below 0.5*M*, since auditors may frequently encounter populations with smaller total error amounts. Third, the effect of rectification on the risks of incorrect rejection should be evaluated for other bounds for the total error amount.

References

American Institute of Certified Public Accountants. 1988. *AICPA Professional Standards*. Vol. 1. Chicago: Commerce Clearing House.

Arens, A. A., and J. K. Loebbecke. 1988. *Auditing: An Integrated Approach*. 4th ed. Englewood Cliffs, NJ: Prentice-Hall.

Cochran, W. G. 1977. *Sampling Techniques*. 3d ed. New York: John Wiley.

Duke, G. L., R. A. Leitch, and J. Neter. 1985. *Behavior of Test Statistics in the Auditing Environment: An Empirical Study*. Sarasota, FL: American Accounting Association.

Garstka, S. J. 1982. The behavior of selected upper bounds of monetary error using pps sampling: A discussion. In *Symposium on Auditing Research IV*. Department of Accountancy, University of Illinois at Urbana-Champaign, 379–85.

Guy, D. M., and D. R. Carmichael. 1986. *Audit Sampling: An Introduction to Statistical Sampling in Auditing*. 2d ed. New York: John Wiley.

Johnson, J. R., R. A. Leitch, and J. Neter. 1981. Characteristics of errors in accounts receivable and inventory audits. *The Accounting Review* 56 (April): 270–93.

Leslie, D. J., and S. J. Aldersley. 1982. The behavior of selected upper bounds of monetary error using pps sampling: A discussion. In *Symposium on Auditing Research IV*. Department of Accountancy, University of Illinois at Urbana-Champaign, 387–99.

———, A. D. Teitlebaum, and R. J. Anderson. 1979. *Dollar-Unit Sampling: A Practical Guide for Auditors*. Toronto: Copp, Clarke, Pittman.

Neter, J., and J. K. Loebbecke. 1975. *Behavior of Major Statistical Estimators in Sampling Accounting Populations—An Empirical Study*. New York: AICPA.

Phillips, J. J. 1985. Bayesian bounds for monetary-unit sampling using actual accounting populations. Ph.D. dissertation, University of Georgia.

Plante, R., J. Neter, and R. A. Leitch. 1985. Comparative performance of multinomial, cell, and Stringer bounds. *Auditing* 5 (Fall): 40–56.

Royall, R. M. 1970. On finite population sampling theory under certain linear regression models. *Biometrika* 57 (2): 377–87.

Smith, T. M. F. 1979. Statistical sampling in auditing: A statistician's viewpoint. *The Statistician* 28 (4): 267–80.

Wurst, J. C. 1987. An investigation of the sieve method and of rectification sampling for auditing. Ph.D. dissertation, University of Georgia.

———, J. Neter, and J. Godfrey. 1989. Comparison of sieve sampling with random and cell sampling of monetary units. *The Statistican* 38 (4): 235–49.

SHORTER ARTICLES

1. Alan Roshwalb and Roger L. Wright
   Using Information in Addition to Book Value in Sample Designs for Inventory Cost Estimation
   (Submitted: April 1989. Accepted: November 1990.)

2. Richard K. Fleischman and Lee D. Parker
   British Entrepreneurs and Pre-Industrial Revolution Evidence of Cost Management
   (Submitted: March 1989. Accepted: November 1990.)

3. Dale Morse, Jens Stephan, and Earl K. Stice
   Earnings Announcements and the Convergence (or Divergence) of Beliefs
   (Submitted: March 1988. Accepted: October 1990.)

4. Bipin B. Ajinkya, Rowland K. Atiase, and Michael J. Gift
   Volume of Trading and the Dispersion in Financial Analysts' Earnings Forecasts
   (Submitted: September 1989. Accepted November 1990.)

5. Scott E. Stickel
   Common Stock Returns Surrounding Earnings Forecast Revisions: More Puzzling Evidence
   (Submitted: June 1989. Accepted October 1990.)

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 348–360

Using Information in Addition to Book Value in Sample Designs for Inventory Cost Estimation

Alan Roshwalb
Georgetown University
Roger L. Wright
RLW Analytics, Cambridge, Massachusetts

SYNOPSIS: Stratified sampling for inventories, as discussed in the literature, usually employs the recorded book value as the stratification variable (Neter and Loebbecke 1975; Roshwalb et al. 1987). Stratification by book value along with an appropriate estimator provides a cost effective and precise estimate of the total inventory cost. Other stratification variables, such as volume of activity, provide information not provided by the book value about the variability of the errors in inventories, but little is known about the effectiveness of designs based on them. Greater improvement in efficiency may be possible by simultaneously including book value and volume into one sample design. In this paper, the effectiveness of volume is investigated as a stratification variable. A bivariate stratification method based on a model-based sampling framework is developed which effectively folds both book value and volume into one collective and simple stratification scheme.

Using a bivariate sample design is not without any cost, because it requires that both the book and the volume values for an item to be positive. However, another extension of the model is proposed to include all items with zero book value or zero volume into a single stratified design.

To examine the effectiveness, sample designs based on book value alone, volume alone, and the bivariate method are constructed for three inventory populations. Since the errors are known for every item in these inventories, the exact efficiency of a sample design can be determined. By designing a sample to achieve a specified precision with known reliability,

The authors wish to thank Jim Godfrey and two anonymous referees for their helpful comments and suggestions on earlier versions of this paper. Financial support was provided in part by the Georgetown University's Summer Research Grant Program and Deloitte & Touche.

Submitted April 1989.
Revisions received January and October 1990.
Accepted November 1990.

the resulting sample size is a measure of efficiency. The design for a population with the smallest sample size is the most efficient and cost effective. To test whether the sampling methodologies actually achieve their precision and reliability targets, the audit sampling process is repeated in a simulation. Random samples are repeatedly selected from the inventories using the alternative sample designs, estimates for the inventory total are calculated using each set of sample data, and the accuracy of the estimates are recorded.

Designs based on the bivariate stratification method required up to 62.5 percent fewer items to achieve the same level of precision and reliability than designs based on book value alone or volume alone. The simulation results show that these bivariate stratified designs attain the same target level of precision and reliability as designs based on book value alone or volume alone. In some cases, the bivariate designs improve the actual confidence interval coverage more than designs based on book value alone.

Key Words: *Audit sampling, Bivariate stratification, Model-based sampling, Sample design.*

Data Availability: *The data used in this study are available from the first author.*

STRATIFIED designs in auditing should be efficient if the sources of errors are related to the stratification variable; there is a strong association between the variability of the errors and stratification variable; and the error rate is moderate to high along with other conditions discussed by Arens and Loebbecke (1981). Although book value is commonly used as the stratification variable, Johnson et al. (1981) considers the following reasons for errors in inventory populations. (1) Errors can accumulate in the inventory management system over the course of a year without detection. (2) Errors in accounts receivable records are often detected by customers, while inventories do not necessarily have a feedback mechanism. (3) Substitutions and returns may not be tracked effectively by the inventory management system. (4) An inventory record may represent many transactions. (5) The cost of cleaning up inventory records may be prohibitive except at year end. Each of these reasons or their effect is arguably related to an item's volume of activity. Thus, stratification based on volume should control the variability and predict the frequency of errors; and a stratified design based on volume should be very effective.

Following a suggestion of Ko et al. (1987) that a model-based stratified design should become more efficient as the complexity of the model's functional form is increased, we include both book value and volume in the model. Also, asymptotic results from Bethel (1989) suggest that, as long as both book value and volume are related to the variability and frequency of the errors, a bivariate stratified design is at least as efficient as other stratified designs.

The remainder of the paper is organized as follows. In the next section stratification methods for one variable are discussed, and the model-based stratification approach is extended to include more than one stratification variable. In section II, the effectiveness of the different sample designs are examined using three inventories. The final sections discuss the extension to include items normally set aside for special handling and brief concluding remarks.

I. Sampling Plans for Inventories

Difference, mean per unit, ratio or regression estimators have been used to estimate the total value of an inventory, and the properties of each have been studied for inventory cost estimation (Baker and Copeland 1979; Beck 1980; Kaplan 1973; Neter and Loebbecke 1975). In this paper, only the difference estimator is considered because of its convenience since the exact forms of the mean and standard error are known.

For an inventory population of N items, let $i=1,2,\ldots,N$ be noninformative labels and define:

X_i = the recorded book value for item i;
Y_i = the audit (line) value for item i;
$D_i=Y_i-X_i$ be the error for item i;

$X=\sum_{i=1}^{N} X_i$, be the total book value;

$Y=\sum_{i=1}^{N} Y_i$, be the total audit value; and

$D=\sum_{i=1}^{N} D_i$, be the total error.

Also, let $j=1,2,\ldots,n$ represent items in the sample, and define: x_j = the recorded book value for item j; y_j = the audit value for item j; and $d_j=y_j-x_j$, the error for item j.

The total value of the inventory, Y, can be estimated using the sample information x_j and y_j and the total book value, X. The difference estimator evolves from the relationship,

$$Y=X+D. \tag{1}$$

The estimator for the total value of a population has the form:

$$\hat{Y}=X+\hat{D}, \tag{2}$$

where $\hat{D}$ is the stratified estimator for the total error:

$$\hat{D}=\sum_{h=1}^{H} N_h \bar{d}_h, \tag{3}$$

with variance:

$$\sum_{h=1}^{H} N_h^2(1-n_h/N_h)S_h^2/n_h, \tag{4}$$

for any stratified design. Here, H is the total number of strata and S_h^2 is the stratum variance such that:

$$S_h^2=\sum_{i=1}^{N_h}\frac{(D_{hi}-\bar{D}_h)^2}{N_h-1},$$

and n_h, N_h, $\bar{d}_h$, and $\bar{D}_h$ are the sample size, stratum size, sample mean of the errors and population mean of the errors for stratum h, respectively (Cochran 1977).

Sample Designs Based on One Design Variable Alone

Sample designs can be divided into probability proportional to recorded book value sampling (PPS) and stratified random sampling (STRS). Although the discussion in this section focuses on sample designs based on book values alone, volume can be substituted for book value anywhere to obtain a sample design based on volume alone.

In PPS sampling, items are selected into the sample with probabilities directly proportional to their recorded book values. In this analysis we focus on STRS sample designs because they have fewer methodological problems (Cassel et al. 1977).

In all STRS methods, separate independent simple random samples are taken from predetermined strata. The sample design specifies the number of strata, the stratum boundaries, and the sample size to be selected in each stratum.

In inventory cost estimation, the stratum boundaries are usually determined using the recorded book values. The efficiency of any STRS plan is affected by the choice of stratum boundaries and the allocation of the sample to the strata. Three methods for determining stratum boundaries are discussed here. They are equal aggregate size (EAS), equal square root (ESR), and model-based stratification (MBSS). The samples are allocated equally among the strata for EAS and MBSS and in proportion to the population size and the standard deviation of the book values within strata for ESR (Arens and Loebbecke 1981; Godfrey et al. 1984). Each of these methods tends to assign higher probabilities of inclusion to items with larger book values.

ESR stratification is more complicated than EAS, the latter develops strata with equal total book values from a list of book values in ascending order. ESR determines the strata by accumulating the square roots of the frequencies of book values (Arens and Loebbecke 1981; Cochran 1961, 1977; Dalenius and Hodges 1959).

MBSS stratification is the most recently developed method (Godfrey et al. 1984; Roshwalb et al. 1987). MBSS stratification relies on a general model:

$$Y_i = \mu + \beta X_i + \epsilon_i, \tag{5}$$

where ϵ_i is a residual error with mean zero and standard deviation:

$$\sigma_i = \sigma_0 X_i^{\gamma}, \tag{6}$$

for inventory item $i=1,\ldots,N$. When $\gamma \neq 0$, the data are heteroscedastic with the model's standard deviation, σ_i, being proportional to X_i^{γ}.

The parameter, β, is assumed to be one and equation (5) can be rewritten as:

$$D_i = Y_i - X_i = \mu + \epsilon_i. \tag{7}$$

The choice of this model is data dependent as indicated in Roshwalb et al. (1987). In some situations, β may not be one; therefore, β should be estimated along with γ, and the appropriate estimator for the total value of a population, Y, is either the ratio or the regression estimator (Baker and Copeland 1979).

The MBSS stratum boundaries are determined by sorting an inventory's items in ascending order of σ_i and finding the stratum boundaries that separate them into H groups with equal within-group sums of σ_i. These boundaries are determined by the specification of σ_i and in the case of equation (6) by the parameter γ; the constant of proportionality, σ_o, is unimportant in this respect. When γ is unknown, the model in equations (6) and (7) must be estimated using methods such as those discussed in Harvey (1976) and Roshwalb (1987).

Sample Design Based on Book Value and Volume Together

The MBSS model can be expanded to allow more than one stratification variable into the design. Book value and volume can be included simultaneously in the equation for σ_i. Instead of specifying the model's residual standard deviation to be equation (6), the multiplicative model is expanded to be:

$$\sigma_i = \sigma_0 (X_{Bi})^{\gamma_1} (X_{Vi})^{\gamma_2}. \tag{8}$$

Here, both the book value, X_{Bi}, and the volume value, X_{Vi}, are assumed to be positive.

The model in equation (8) reduces two variables to a single variable σ_i, and again, using a list of items in ascending order of σ_i, the MBSS stratum boundaries are the points where the sum of σ_is are the same for each stratum. Thus, the specification of σ_i is the only difference between the univariate and bivariate methods of MBSS stratification. As in the univariate situation, the values of γ_1 and γ_2 are either assumed or estimated. The algorithms in Harvey (1976) and Roshwalb (1987) provide for the simultaneous estimation of γ_1 and γ_2.

II. Empirical Findings

Roshwalb et al. (1987) studied EAS, ESR, and MBSS stratified sampling plans based on book values for four inventory populations where the book values and audit values are known for all items in the population. Of these inventories, two have volume variables available and are used in this study. In one inventory (referred to as RPD for the inventory of a retail pharmaceutical firm), volume is measured by the total value shipped for each item. The second inventory (referred to as IMD1 for a finished goods warehouse of an industrial manufacturing plant), volume is the dollar value of the previous year's actual sales plus returns for each item.[1]

An added third data set, IMD2, examines the same warehouse as IMD1 but for another year. Due to changes in warehousing policy, IMD2 contains far more items than IMD1, which allows treating them as two distinct inventories.

Description of the Study Populations

Unlike accounts receivables and some other accounting populations, inventory populations have much higher error rates and a high incidence of both under- and overstatement errors (Johnson et al. 1981). Although the three inventories examined here have higher error rates than those reported in Johnson et al. (1981), they have similar error means, standard deviations, and skewness measures.

The sizes of our inventories differ (table 1); RPD has 8,232 items, IMD1 has 5,103 items, and IMD2 has 11,607 items. Each of the inventories contains a substantial number of items with zero book values, but more than 90 percent of these items have positive volume. Also, a small number of items have had no activity in the past year while maintaining a positive balance.

[1] Actual sales alone has also been examined, but the results were nearly identical to the results using sales plus returns and are not reported here.

Table 1
Population Sizes

Inventory	Total Database Size	Number of Items[1]: Positive Book and Volume	Positive Book and Zero Volume	Zero Book and Positive Volume	Zero Book and Volume
RPD	8,232	7,770	299	163	0
IMD1	5,103	3,723	38	1,263	79
IMD2	11,607	3,962	33	6,948	664

[1] The number of items with positive book value and positive volume along with positive book and zero book is the same as reported for RPD in Roshwalb et al. (1987). For IMD1, the volume measure and the book values and errors came from two different administrative departments' databases. Any discrepancy reflects non-matches in the records of the two departments.

Table 2
Percentage of Errors in Population

Inventory	N	Overstatement Error	Understatement Error	Overall Error Rate
RPD	7,770	35.3	48.4	83.7
IMD1	3,723	50.1	41.8	91.9
IMD2	3,962	67.8	25.9	93.7

The distributional characteristics of the inventories, reported in tables 2 and 3, show a substantial number of both over- and understatement errors, high error rates, and highly variable and kurtotic errors in all cases.[2]

The increasing variability of the errors with increasing book values shown in Roshwalb et al. (1987) for RPD and IMD1 is also observed for IMD2. The errors also exhibit increasing variability for increasing volume. Given these relationships, we expect that a stratification plan based on either variable will effectively control the variability of errors.

Roshwalb et al. (1987) also show that the log-linear relationship:

$$\log(\sigma_i)=\log(\sigma_o)+\gamma_1\log(X_{Bi}), \tag{9}$$

suggested by equation (6) with book value as X_i, exists. The comparable log-linear relationship for volume is:

$$\log \sigma_i=\log \sigma_o+\gamma_2\log(X_{Vi}). \tag{10}$$

---

[2] Setting aside the items with zero book value or zero volume reduces the total number of items available for the analysis. Tables 1, 2, and 3 reflect these reductions and report different values than Roshwalb et al. (1987).

Table 3
Distribution Characteristics of Book Value, Error, and Volume

Inventory	*Variable*	N	*Total* ($)	*Mean* ($)	*Standard Deviation* ($)	*Skewness*	*Kurtosis*	*Minimum*	*Maximum*
RPD	Book	7,770	7,805,625	1,004.6	3,093	12	214	0.01	8,458
	Error	7,770	55,594	7.2	263	−9	379	−7,948.80	4,925
	Volume	7,770	66,563,487	8,566.7	21,229	11	181	0.01	525,860
IMD1	Book	3,723	137,240,000	36,830.0	147,270	23	815	1.56	612,969
	Error	3,723	495,660	−130.3	23,076	6	258	−400,600.0	551,676
	Volume	3,723	727,500,000	195,420.0	1,138,194	15	275	0.72	29,358,000
IMD2	Book	3,962	166,043,458	41,909.0	239,740	31	1,334	2.55	11,434,000
	Error	3,962	−172,132	−434.4	38,530	6	283	−742,700.0	1,087,700
	Volume	3,962	791,211,400	199,700.0	1,711,800	30	1,252	0.07	80,085,000

Table 4
Estimates of Model Parameters Used in MBSS Sample Designs

Population	*Variables in Model*	μ	σ_0	γ_1 *(Coefficient for Book Value)*	γ_2 *(Coefficient for Volume)*
RPD	Book alone	3.6	2.8	0.5582	—
	Volume alone	−1.1	30.3	—	0.4656
	Book and volume	0.1	1.6	0.4387	0.1925
IMD1	Book alone	382.7	106.3	0.3744	—
	Volume alone	186.7	549.7	—	0.3500
	Book and volume	−9.8	12.4	0.3887	0.3001
IMD2	Book alone	−433.9	313.5	0.4340	—
	Volume alone	56.9	143.6	—	0.4326
	Book and volume	4.4	17.6	0.2683	0.3687

Model estimates of equations (9), (10), and the bivariate model in equation (8) are presented in table 4. The associated model would not be justified in any of these cases, if γ_1 or γ_2 is close to zero. In all cases, both the univariate and bivariate models are justified by the estimates of γ_1 and γ_2.

Sample Design Results

In this paper, 20 strata are used.[3] To observe the effectiveness of the different stratification methods, we will calculate, using the variance expression in equation (4), the sample size required to achieve a precision of two percent the total book value with 95 percent reliability. The stratum standard deviations are computed using all the errors.[4] The relative difference between the two designs' sample sizes measures the relative efficiency of any two plans for one inventory.

The required sample sizes for the three inventories under various sampling plans are given in table 5. Based on book value alone, ESR designs are more efficient than EAS designs. EAS designs require from 0.7 percent to 24.2 percent more observations than an ESR design. MBSS designs are more efficient than an ESR design because the former require from 7.3 percent to 15.7 percent fewer observations.

ESR or MBSS designs based on volume alone perform better than ESR or MBSS based on book value alone for IMD1 and IMD2, the savings range from 35.6 percent to 55.1 percent. However, the plans based on volume for RPD are less efficient than the ESR designs based on book value.

The bivariate MBSS sampling plans based on the combined book value and volume

---

[3] The number of strata in any design is at an auditor's discretion. It is not considered unusual to use up to 20 strata in accounting applications (Arens and Loebbecke 1981). The gains from using more strata are not significant when going from 15 to 20 strata (Neter and Loebbecke 1975). However, both the ESR and MBSS methods achieve certain optimality properties when the number of strata are large (Bethel 1989).

[4] Usually, the ESR method calculates the required sample size using stratum standard deviations of the book values. After sample selection, the precision of the estimator is estimated using the stratum standard deviations of the errors. By using the stratum standard deviations of the errors in the sample size calculation, the sample size is a measure of the estimator's efficiency.

Table 5

Required Sample Sizes and Relative Efficiencies for 20 Strata Designs

	Stratified Plan						
	Book Value Alone			*Volume Alone*			*Book and Volume*
Inventory	*EAS*	*ESR*	*MBSS*	*EAS*	*ESR*	*MBSS*	*MBSS*
RPD	264 (−21.1)	218 (0.0)	201 (7.8)	377 (−72.9)	285 (−30.7)	303 (−39.0)	191 (12.4)
IMD1	1,475 (−24.2)	1,188 (0.0)	1,002 (15.7)	1,319 (−11.0)	765 (35.6)	699 (41.2)	630 (47.0)
IMD2	1,141 (−0.7)	1,133 (0.0)	1,050 (7.3)	989 (12.7)	516 (54.6)	509 (55.1)	425 (62.5)

Note: ESR design based on book value is used as the basis for comparison and the reported relative efficiency is 0.0.

require substantially fewer observations than the designs based on book value alone. The savings range from 12.4 percent to 62.5 percent.[5]

Based on the sample sizes reported in table 5, the precision and reliability rates are investigated for univariate ESR, univariate MBSS, and bivariate MBSS designs using a simulation of 2,000 replications. Each design was constructed to yield estimates that would fall within two percent of the total book value with 95 percent reliability. From table 6, all designs are close to achieving this level of precision. The designs based on book value alone for IMD1 and IMD2 have lower coverage rates than the nominal 95 percent due to the kurtotic nature of the data. Designs using volume alone or the bivariate model control the kurtotic nature of the data, and the confidence interval coverage rates are close to the nominal 95 percent rate.

From the simulation, estimates from the bivariate MBSS method are just as reliable and precise as estimates from any other design but for smaller sample sizes. An explanation for some of the improvement can be seen in table 7, which shows 20 strata designs for IMD1 using ESR and MBSS based on book value alone and bivariate MBSS. The plans are broken down by strata into relative sampling fractions (i.e., $(n_k/N_k)/(n/N)$) and standard deviations. The strata sampling fractions for ESR and MBSS are increasing for the strata with the larger book values. The high sampling fraction tends to reduce the contribution of a stratum's high variance to the variance of the difference estimator in equation (4). However, the strata with the smaller book values have unexpectedly large standard deviations. In particular, the standard deviations for stratum 2 of the MBSS plan and stratum 3 of the ESR plan are very large relative to their

---

[5] The only noticeable characteristic which affects a sampling plan based on volume's effectiveness is the association between the errors and volume. For RPD this relationship does not appear to be as strong as with the other populations. For example, the error standard deviations are 859.59 for items with zero volume and positive book, 329.49 for items with positive volume and zero book, and 263.31 for items with positive book value and positive volume. In each of these categories volume is increasing, therefore, the standard deviations should also be increasing as observed for IMD1 in table 9. It seems that book value may sometimes have a stronger relationship to the variability of the errors than volume and provide a more efficient univariate sample design, but volume may still contribute to explaining the variability through the bivariate model.

Table 6
Precision and Reliability Rates from Simulation

Stratification Variable	*Book*		*Volume*		*Book and Volume*
Sample Design	*MBSS*	*ESR*	*MBSS*	*ESR*	*MBSS*
Point Estimates within 2%					
RPD	95.0	95.6	95.4	95.0	95.4
IMD1	94.6	93.2	95.2	95.2	93.2
IMD2	96.8	97.4	95.8	95.4	95.5
Confidence Interval Coverage Rate					
RPD	95.1	94.9	94.1	95.6	95.5
IMD1	81.7	74.2	94.0	89.7	91.6
IMD2	84.5	81.6	95.1	95.8	95.0

Note: Simulation consists of 2,000 samples constructed using the knowledge of the entire population.

Table 7
Relative Sampling Fractions and Within-Stratum Standard Deviations for 20 Strata Designs

	MBSS with Book Value Alone		*ESR*		*MBSS with Book Value and Volume*	
Stratum	*Relative Sampling Fraction*	*Error Standard Deviation*	*Relative Sampling Fraction*	*Error Standard Deviation*	*Relative Sampling Fraction*	*Error Standard Deviation*
1	0.28	14,139	0.38	13,659	0.18	1,932
2	0.50	49,261	0.47	28,075	0.41	1,534
3	0.62	16,844	0.57	46,615	0.53	2,822
4	0.72	11,786	0.60	17,461	0.67	3,111
5	0.80	19,050	0.73	13,350	0.80	8,906
6	0.89	13,553	0.75	14,689	0.93	4,392
7	0.97	2,977	0.86	13,845	1.07	5,102
8	1.05	9,490	0.95	6,478	1.23	6,320
9	1.12	10,758	1.03	12,121	1.43	5,370
10	1.23	7,018	1.31	5,704	1.65	16,365
11	1.34	7,381	1.37	7,777	1.89	18,885
12	1.42	4,599	1.53	7,396	2.10	11,175
13	1.52	3,300	1.67	5,672	2.36	11,983
14	1.65	18,877	1.73	2,992	2.64	30,533
15	1.81	49,329	1.98	3,512	3.06	22,781
16	2.00	13,642	2.05	12,224	3.56	20.617
17	2.23	2,174	2.49	20,462	4.03	21,731
18	2.44	2,768	3.46	72,621	4.50	29,785
19	2.80	72,405	2.77	13,759	5.47	81,140
20	3.14	4,571	3.25	15,134	6.12	28,054

Note: The relative sampling fraction is the ratio of the stratum sampling fraction to the overall sampling fraction. It measures the relative sampling weight assigned that stratum.

neighboring strata. The relatively small sampling fractions in these strata inflate the contributions of the large variances to the variance of the estimator. Bivariate MBSS sampling determines stratum membership by book value and volume together. The items assigned by the first two designs to strata for small book value are now assigned to other strata with more appropriate sampling fractions. Consequently, there is a more suitable correspondence between the sampling fraction and the standard deviations, yielding a more efficient sample design.

Unlike this setting, the complete population is not known in practice. However, a planning sample may be available, and the information from such a sample can be used in lieu of population data.

III. MBSS Sampling Plan That Includes All Inventory Items

A simple adaptation to the model is proposed so that the MBSS approach can encompass items usually set aside for special handling such as those having zero book values but positive volume. Instead of specifying the model for the residual standard deviation to be expression (8), let the model be:

$$\sigma_i = \sigma_o(1+X_{Bi})^{\gamma_1}(1+X_{Vi})^{\gamma_2}. \tag{11}$$

Other transformations may be more effective than equation (11), but the effect of adding one to book value and volume for use in determining σ_i is relatively small. This transformation also assures that the value of σ_i is always greater than zero.

Using equation (11), MBSS strata can be constructed such that all items can be included in a unified design, which can be illustrated by examining a 20-strata bivariate MBSS design for IMD1. The sampling plan uses the following estimated values:

$$\hat{\gamma}_1 = 0.3813,\ \hat{\gamma}_2 = 0.3001, \text{ and } \hat{\sigma}_o = 2{,}557.$$

Table 8 provides the strata boundaries, sizes, and relative sampling fractions for each stratum. Strata 1 has 2,430 members of which the largest value for $\hat{\sigma}_i$ is 480,052. Stratum 2 has 477 items. Any item is a member of stratum 2 if $\hat{\sigma}_i$ is between 480,052 and 744,472.8.

For an item with zero book and a positive volume, the stratum of the item is determined by the size of the volume variable. In table 9, the minimum value of volume with zero book value is \$0.34. For this item, $\hat{\sigma}_i = 2{,}791.7$, and the item belongs to strata 1. The maximum value of volume with zero book value is \$521,370. For this item, $\hat{\sigma}_i =$ 132,876, and this item belongs to strata 5. The relative sampling fraction for the first item is 0.11 and for the second item is 1.15. That is, the probability of selection is over ten times greater for the second item than the first.

Comparing this unified MBSS design to a MBSS plan that excludes items with zero book value or zero volume together with a simple random sample (SRS) of the excluded items, we find the unified MBSS design to be about two percent more efficient. Although the unified MBSS design is only moderately more efficient than the MBSS plus SRS design, it simplifies the analysis of the inventory considerably since only one plan is required.

IV. Discussion

The results in this paper indicate that by including book value and volume into a bivariate model, the efficiency and performance of stratified designs can be improved

Table 8

Unified 20-Strata MBSS Design

Strata	Maximum Value of $\hat{\sigma}_i$	Stratum Size	Relative Sampling Fraction
1	480,052.0	2,430	0.10
2	744,472.8	477	0.52
3	981,052.3	341	0.72
4	1,228,811.1	265	0.93
5	1,510,023.5	214	1.15
6	1,743,793.4	180	1.37
7	2,030,626.2	155	1.59
8	2,372,043.9	133	1.85
9	2,769,840.3	114	2.16
10	3,139,270.9	99	2.49
11	3,547,466.9	87	2.84
12	4,039,174.8	78	3.16
13	4,593,703.9	68	3.63
14	5,187,388.7	59	4.18
15	5,927,770.4	52	4.74
16	6,847,986.4	46	5.36
17	7,655,690.7	40	6.17
18	8,712,492.9	36	6.85
19	9,703,287.5	32	7.71
20	11,010,160.3	28	8.81

Table 9

IMD1 Population Error, Book, and Volume Totals for Nonpositive Book and Volume Groups

Subgroup	N	Variable	Total	Mean	Standard Deviation	Minimum	Maximum
Positive book and zero volume	38	Error	191,145	77.49	2,532	−10,029.00	10,725
		Book	188,197	4,952.60	10,908	7.59	47,565
Zero book and positive volume	1,263	Error	707,076	559.84	3,884	0.00	59,086
		Volume	54,973,181	43,526.00	196,720	0.34	521,370
Zero book and zero volume	79	Error	162,846	2,061.30	18,209	0.00	161,860

over univariate book value designs. The improvement in efficiency can be expected when the error rate is moderate to high, over- and understatement errors are present, and errors and their causes are associated with a measure of activity such as volume.

The results in this study are based on three inventory populations. The ability to generalize beyond these three data sets is limited. As often mentioned, a serious limitation to any study of statistical estimators and sampling methods is the lack of standardized data sets for comparative analysis (Menzefricke and Smieliauskas, 1987). Although the mean and standard deviation of the errors in our inventories are similar to those other studies, the error rates, 83.7 percent to 93.7 percent, are higher than other previously reported data sets, a maximum of 75.8 percent in Johnson et al. (1981). Additional data sets with different error rates along with volume measures are required to study any effect on efficiency due to changes in error rates.

Also the analysis is based on the knowledge of all the errors, the results do not contain variability due to sample selection and are considered the expected value of a design's precision estimated from a representative random sample. If the procedures are to be used in practice, a planning sample is required to estimate the parameters of the model. In this case, the efficiency of the sample design relies on the representativeness of the sample.

References

Arens, A. A., and J. K. Loebbecke. 1981. *Applications of Statistical Sampling to Auditing.* Englewood Cliffs, NJ: Prentice-Hall.

Baker, R. L., and R. M. Copeland. 1979. Evaluation of the stratified regression estimator of auditing accounting populations. *Journal of Accounting Research* 17 (Autumn): 606–17.

Beck, P. 1980. A critical analysis of the regression estimator in audit sampling. *Journal of Accounting Research* 18 (Spring): 16–37.

Bethel, J. 1989. Minimum variance estimation in stratified sampling. *Journal of the American Statistical Association* 84 (March): 260–65.

Cassel, C. M., C. E. Sarndal, and J. H. Wretman. 1977. *Foundations of Inference in Survey Sampling.* New York, NY: Wiley.

Cochran, W. G. 1961. Comparison of methods determining stratum boundaries. *Bulletin of the International Statistical Institute* 38: 345–58.

———. 1977. *Sampling Techniques.* New York, NY: Wiley.

Dalenius, T., and J. L. Hodges. 1959. Minimum variance stratification. *Journal of the American Statistical Association* 54 (March): 88–101.

Godfrey, J., A. Roshwalb, and R. L. Wright. 1984. Model-based stratification in inventory cost estimation. *Journal of Business and Economic Statistics* 2 (January): 1–9.

Ham, J., D. Losell, and W. Smieliauskas. 1985. An empirical study of error characteristics in accounting populations. *The Accounting Review* 60 (July): 387–406.

Harvey, A. C. 1976. Estimating regression models with multiplicative heteroscedasticity. *Econometrica* 44 (May): 461–65.

Johnson, J. R., R. A. Leitch, and J. Neter. 1981. Characteristics of errors in accounts receivables and inventory audits. *The Accounting Review* 56 (April): 270–93.

Kaplan, R. S. 1973. Statistical sampling in auditing with auxiliary information estimators. *Journal of Accounting Research* 11 (Autumn): 238–58.

Ko, C., C. J. Nachtsheim, G. L. Duke, and A. D. Bailey. 1988. On the robustness of model-based sampling in auditing. *Auditing: A Journal of Practice & Theory* 7 (Spring): 119–36.

Menzefricke, U., and W. Smieliauskas. 1987. A survey of simulation studies in statistical auditing. *Journal of Accounting Literature* 6: 26–54.

Neter, J., and J. K. Loebbecke. 1975. *Behavior of Major Statistical Estimators in Sampling Accounting Populations: An Empirical Study.* New York, NY: American Institute of Certified Public Accountants.

Roshwalb, A. 1987. The estimation of a heteroscedastic linear model using inventory data. *American Statistical Association: Proceedings of the Business and Economics Section* 321–26.

———, R. L. Wright, and J. Godfrey. 1987. A new approach for stratified sampling in inventory cost estimation. *Auditing: A Journal of Practice & Theory* 7 (Fall): 54–70.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 361–375

British Entrepreneurs and Pre-Industrial Revolution Evidence of Cost Management

Richard K. Fleischman
John Carroll University
Lee D. Parker
The Flinders University of South Australia

SYNOPSIS: Accounting histories have dated the advent of sophisticated cost management from the mid-1880s (Solomons 1952). The scientific management movement is credited with instituting and popularizing cost management techniques. However, it might be suspected that British entrepreneurs of the Industrial Revolution would have developed sophisticated costing techniques earlier, given their significant methodological advances in other economic areas.

This article reports the findings from surviving business records of 25 sizeable British industrial firms (mostly in the iron and textile industries) from 1760 to 1850. Substantial evidence of a relatively mature cost management has been found in four major areas of activity: cost control techniques, accounting for overhead, costing for routine and special decision making, and standard costing. Speculations about the motivations for cost management and about specific factors influencing the iron and textile industries are considered. Because the accounting practices of these firms predated the genesis of "the costing renaissance" a century later, our understanding of cost management practices in the Industrial Revolution is augmented by the survey.

Submitted March 1989.
Revisions received September 1989, March and October 1990.
Accepted November 1990.

Key Words: *Accounting history, Cost-accounting, Cost management, Industrial Revolution.*

Data Availability: *Data used in this study may be obtained from the authors on request.*

I. Historians' Views of Industrial Revolution Costing

POLLARD has been the historian most cited on the subject of Industrial Revolution management. Having examined existing business records, he concluded that "the practice of using accounts as direct aids to management was not one of the achievements of the British Industrial Revolution" (Pollard 1965, 248). While conceding that accounting addressed issues such as stock control, embezzlement, and departmental comparisons, he argued that it did not guide management decisions.

Observers have regarded the Industrial Revolution's cost-accounting as unsophisticated for a variety of reasons. Fixed costs were not allocated, either because the amounts involved were minimal or because the relationship between indirect expenses and product costing was not understood (Chatfield 1977; Edwards 1937; Pollard 1965). Wells (1978) contended that sophisticated costing could not have prevailed until the post-Industrial Revolution increase in fixed assets necessitated a more careful allocation of common costs. Johnson (1972) and Chatfield (1977) argued that a proper integration of cost and commercial accounts was not accomplished until business firms became sufficiently complex to require material cost transfers.

These views of Industrial Revolution costing may partly be explained by the absence of an eighteenth-century industrial-accounting literature. Accounting textbooks dealt almost exclusively with merchant-accounting techniques. Babbage's (1835) was the only cost-accounting text of note during the Industrial Revolution. Garcke and Fell's subsequent first edition (1887) is most cited as the genesis of modern cost-accounting theory (Johnson 1972; Parker 1986). This literary shortfall is said to reflect entrepreneurs' inattention to production costs since profit margins were considered high (Edwards 1937; Hudson 1977; Parker 1986; Pollard 1965). It has also been suggested that the orientation of entrepreneurs toward technological innovation distracted them from the more commercial side of their firms (Parker 1969; Urwick and Brech 1964). Thus, Industrial Revolution cost-accounting has largely been deemed unsophisticated even though the larger enterprises of the period were substantial in terms of capital investment, personnel, and process integration. Structural features of accounting during the period have reinforced these views: British accounting education lagged (Stacey 1954); the few accountants in public practice rarely serviced manufacturers (Parker 1986); and there were no accepted accounting conventions (Yamey 1960).

Several previous historical pieces have suggested more sophisticated Industrial Revolution cost management. Roll (1968) found evidence of superior cost management in his study of Boulton & Watt's Soho foundry. He showed evidence of time and motion data, elaborate costing calculations (including overhead), scientific pricing, departmental profit analysis, and process efficiency testing. McKendrick (1970) found that the Wedgwood company employed intricate cost data for determining pricing policy and analyzing the economics of large-scale production. Wedgwood was also concerned that the output of certain productive lines be sufficient to justify the fixed costs. Research in early nineteenth-century textile mills has produced additional fruitful results. At the Charlton Mills, Stone (1973, 78) uncovered cost-accounting techniques "not believed to have been in use until the early twentieth century," including the allocation of overhead to cost centers by means of predetermined rates and rudimentary transfer pricing.

II. Research Methodology

The authors have examined the surviving business records of 25 large British manufacturing enterprises of the Industrial Revolution (see the appendix for data sources). Major manuscript collections utilized by Pollard (1965) and archives used by economic historians in writing industry- and firm-specific studies of the Industrial Revolution period were examined. Large companies whose primary orientations were transport and extraction were generally not studied because of our expectation that nascent cost-accounting would flourish more notably in a factory environment. Only firms located in Scotland, Wales, and the manufacturing districts of England were investigated. The sample is disclosed in table 1.

Evidence proved difficult to uncover because costing activities were rarely reported in financial accounting records. More fruitful, though less plentiful, were letters between owners, notebooks of prominent personnel, company minute books, and production reports. The archives of absentee-owned firms typically featured a flow of costing correspondence between managers and owners. The available records reflect wide individual disparities, both quantitatively and qualitatively. Unfortunately, most records are lost for a number of concerns whose surviving documents suggest the use of sophisticated techniques (e.g., Ashworth, Herculaneum, and Crowley), while substantial holdings exist for some firms almost devoid of cost-accounting practice (e.g., James Finlay, McConnel & Kennedy, Peter Stubs, and Newton, Chambers).

We estimate the survival rate of business records for the majority of Industrial Revolution firms, at least through the 1840s, as less than ten percent. This shortfall has made it extraordinarily difficult to trace actions taken subsequent to costing analyses or projections.

III. Cost Management Practices in Review

Surviving business records attest to a frequent use of certain cost management techniques, which are reviewed here in four categories of cost-accounting activity.

Cost Control

The highly competitive textile and iron industries of the Industrial Revolution routinely employed cost data to control raw material inputs. Textile firms were concerned about profitable raw material mixtures and waste standards. Iron firms were most attentive to differential mining costs, royalty payments, and yields in their various mineral fields. Typical of this raw material control was the notebook of Thomas Beer, the accountant at Mona Mines, who reported data on good copper yields from different ores, taking into account differential costs of mining and smelting.

Marshall, a textile firm, controlled expenses through the maintenance of daily performance cards for each spinning frame, input/output records for every grade of raw material, and semiannual compendiums of all costs associated with their four productive processes. The Gregs, also in textiles, kept wage and expenses books in

Table 1

British Industrial Revolution Firms Reflecting Activity (A) and Excellence (E) in Cost Management Areas

Firms	Industry	*Cost Control (Responsibility Management)*	*Overhead Accounting (Product Costing)*	*Decision Making (Routine and Special)*	*Standard Costing (Budgets, Forecasts Inventory Control)*
Ashworth	Cotton	E	A		E
Boulton & Watt	Machining	E	E	E	E
Cardwell, Birley	Cotton				A
Carron	Iron	E	E	E	E
Crawshay	Iron	A	A		A
Crowley	Iron	E		A	E
Darby	Iron	E	A	E	A
Dowlais	Iron	E	E	E	A
Dugdale	Cotton				A
Finlay (James)	Cotton				A
Gott	Wool	E	A	E	
Greg	Cotton	E	A	A	A
Herculaneum	Pottery	A		A	A
Horrocks	Cotton		A		
Marshall	Flax	E	E	E	A
McConnell & Kennedy	Cotton	A			
Mona Mines	Copper	A	A	A	A
Nasmyth, Caskell	Machining	A			A
Newton, Chambers	Iron				A
Oldknow	Cotton	A	A	A	A
Philips	Tape		A	E	E
Robinson	Cotton	A	A		
Strutt	Cotton	E	E	A	E
Stubs	Steel				
Wedgwood	Pottery	A	A	A	A
Total Activity		18	16	14	20
(Percentage of 25 Firms)		72	64	56	80

which they calculated important control statistics. Losses in 1829 were attributed to heightened production costs associated with using older machinery, better quality cotton than necessary, and superfluous card room overseers.

The most meticulous expense book was found in the Strutt archives. Recorded expenses at the Belper plant included weekly costings of such items as superintending, dinner breaks, and "learning." Moreover, expenses were measured to a thousandth of a pence, with overhead allocated to various processing departments. At the Milford location, expenses per pound of yarn were recorded quarterly by operation and averages were calculated for the nine different process cost centers.

Cost control was also advanced via regular cost comparisons. Mills, departments, blast furnaces, steam engines, other machines, and workers were rated against one another in terms of costs and profitability. Large firms engaged in macro-level cost comparisons, either at the function room or plant level. Quarterly cost comparisons were used to evaluate the cost control of respective managers in the four carding and four spinning rooms of Strutt's mill at Milford. Boulton & Watt compiled weekly expense reports comparing their multiple lathing and drilling rooms. In the 1790s, the blast furnace supervisor at Carron was transmitting monthly reports of costs per ton for good iron for each blast furnace, with notes explaining why production might be down either generally or at a particular furnace (Fleischman and Parker 1990).

In evaluating steam engine efficiency, Gott compared the expenses of working one old and one new engine, taking five-year averages during the period 1833–1857. A Dowlais notebook from 1827–1828 provided intricate, cost-based detail on the horsepower generated and coal consumed by each engine. The weekly accounts for the blast furnaces at Darby likewise tracked and costed the downtime and coal consumption of each engine.

Responsibility cost management was in evidence in a significant minority of firms that attempted to record expenses at the individual cost or profit center level. Gott established their scribbling, dyeing, mechanics, and milling operations as independent profit centers. Darby, in the 1820s, undertook a management reorganization that elevated the best men to managerial rank with responsibility for performance reporting. Separate profit and loss statements were compiled for individual components of the firm and for departments within each segment. From its earliest history in the 1760s, Carron ownership insisted that managers be responsible for the performance and profitability of their own departments.

Accounting for Overhead

There are numerous complex product-cost calculations in surviving business records. More sophisticated efforts included factory overhead items and selling expenses, in addition to the more obvious prime costs.

Boulton & Watt maintained engine books from the early nineteenth century. Hundreds of components comprising finished engines were individually costed. Josiah Wedgwood communicated cost calculations to his partner for 31 different vase lots in 1772. These costings included a wide range of overhead items and period expenses (McKendrick 1970). Boulton & Watt and Wedgwood were in the enviable position of being able to base selling prices on these product costings. Firms in the more highly competitive textile and iron industries were more dependent on a market-driven pricing structure. However, there were sufficient product lines, differential product qualities, and regional variations that accurate costing efforts aided in determining the most profitable output (Edwards 1989).

The Greg manuscripts contain an 1831 calculation that shows the costs associated with producing 18,000 pounds of twist weekly. Included were an allocation to recover the initial outlay for fixed assets, raw materials, waste, wages, and an itemized "contingencies" list. The resulting product cost was 2.27 pence per pound. Marshall kept semiannual accounts for each branch of the flax business in more than 40 data categories. This information was utilized for technology assessment, profitability calculations, yield and production comparisons, and managerial evaluations (Rimmer 1960). Leading iron companies also accounted for overhead cost. In 1763, Carron Company's General Court ordered a breakdown of all costs for each product individually, including applied overhead. In 1798, Darby management observed that many manufactured goods did not have prices proportional to the variety of "additional charges incidental to manufacturing." A resident manager at Crawshay was advised in 1814 not to close down a blast furnace because it would decrease production, increase the common charge allocation per unit of output, and raise prices.

Four large Industrial Revolution firms allocated overhead to departments based on formulae. Carron Company made annual allocations to its seven cost centers, specifying what constituted "general charges" as opposed to specific departmental expenses. Marshall's Shrewsbury accounts contained detailed fixed overhead allocation tables. When the amount of a fixed overhead component changed, the adjustment was not passed on to all cost centers proportionately but only to some, as a function of use. Likewise, at the Dowlais iron and the Strutt textile enterprises, differential apportionments were made to cost centers on a predetermined basis.

Garcke and Fells (1893) claimed that very few factories allocated indirect costs to jobs. However, a substantial number of Industrial Revolution firms included in "general charges" a wide variety of costs integral to project evaluation and job costings. Reliance on predetermined application rates, though not universal, was a procedure of several firms. The methods resembled those suggested by Metcalfe (1885) but lacked the sophistication of techniques advanced by Church (1916).

Costs for Routine and Special Decision Making

Industrial Revolution entrepreneurs employed cost analyses to support vertical integration and product line decisions. Capital equipment purchases, mineral field leases, and major technology decisions (e.g., the introduction of the power loom) were all undertaken after a careful costing of alternatives. Subcontracting and transport options were likewise evaluated.

Two giant iron concerns, Darby and Carron, were clear leaders in generating cost analyses to aid decision making. Darby's management acted on five cost analyses covering a variety of decisions in 1793. In 1798 and 1825, the firm carefully considered the viability of the Coalbrookdale and Horsehay leases. In 1813, the firm did an intricate calculation of the expenses and savings associated with the use of a boiler to conduct heat from the puddling furnaces, a development Raistrick (1970) called a very significant technological advance. In 1768, Carron had a report prepared on the extraction and profitability of Kinnaird coal, which included time and motion studies on the mining process and laid the foundation for standards and transfer prices. In 1770, a survey of coal reserves was made and differential costs were calculated for the mining of each.

Included was a cost comparison between existing extraction levels and a proposed reduction to 50 percent of current volume. Carron's management frequently demanded cost data for items such as subcontracting, machinery procurement, and repair and replacement decisions (Fleischman and Parker 1990).

One aspect of the Industrial Revolution was the growing vertical integration of larger firms, often as a result of cost/profit estimations. Though Crowley concentrated on the manufacture of finished iron products almost exclusively, the firm did expand into pig iron forging and outright ownership of its own shipping fleet (Flinn 1962). Samuel Oldknow made informed decisions as his firm added new processes, such as bleaching, printing, and dyeing (Unwin 1924). Wedgwood considered stove manufacturing, but concluded that the expansion would not be cost effective for the space required (Euphemia 1973).

The abandonment of unprofitable product lines was occasionally done on the basis of costing. In 1765, Carron Company separated and evaluated the blacksmith and plating forge accounts, which resulted in the discontinuance of anchor production. Boulton & Watt manufactured many of their steam engine components, but frequently made and reversed subcontracting decisions for parts (Tann 1981). Wedgwood either abandoned or de-emphasized crests, figurines, and pottery painting (Euphemia 1973).

Transportation costs were a matter of consequence to some firms. Herculaneum imported coals via the Sankey Canal after alternative routes were investigated. Wedgwood did a cost/benefit analysis in 1765 of constructing a canal into Staffordshire (Euphemia 1973). Mona Mines did a cost comparison of selling its output at Swansea rather than Liverpool in 1827.

Standard Costing

Twenty firms in our sample employed production and costing standards. These were frequently utilized for effective inventory control and occasionally for the preparation of budgets and forecasts. At Crowley, an inventory taken every ten weeks became the basis of production plans budgeted by their accountant (Flinn 1957). The Philips archive contains a detailed set of production standards for 1771 that shows monthly input quantities projected over a 20-month period to achieve a desired level of output. Henry Ashworth established standards for various twist counts in his 1849 diary. In each case, Ashworth calculated a "prime cost" per pound on the basis of a spinner's labor standard and a raw materials standard, which included a waste allowance. In one case, selling price was a mere halfpenny in excess of the prime cost standards, indicative of the narrowness of profit margins in cotton textiles.

Boulton & Watt maintained a book of moulders' estimates for orders during the years 1815–1839, with raw materials and labor processes charged at standard. The risk of irreversible damage during the boring process was charged at one-twelfth these costs, tools at 25 percent of job cost, with a bottom line markup of 40 percent. Other firms, like Nasmyth and Wedgwood, also took care in tracking prime costs to specific orders. There is no evidence to suggest that Metcalfe (1885) was aware of these developments when he wrote about job-order costing a half-century later.

In the 1830s, furnacemen at Darby agreed on 60 tons per week as the norm, with bounties paid for excess production and abatements levied below (Trinder 1974). Mona Mines contracted "tribute bargains" with its miners every two months. For each mine and each shaft, a standard price per ton of ore was established. Variances between actual and standard labor costs were placed into a sundry account.

The use of standard costing was associated with an awareness of the cost savings available from appropriate inventory control procedures. Greg and Herculaneum owners were sensitive to stockout dislocation problems. Dowlais and Wedgwood correspondence cautioned against a variety of inventory carrying costs (Elias 1960; Euphemia 1973). Most firms used standard costs for valuing inventory and were alert to the need to eliminate obsolete items.

Boulton & Watt wrote down saleable goods that had decreased in value. Both manufactured cost and net realizable value were recorded in inventory, but only the lower figure was included in the final total. Nasmyth, Gaskell conducted quality checks on raw materials received (Cantrell 1984); Cardwell, Birley tracked and included twist and weft in process; and Darby's Ketley foundry maintained records for the transition of output from work-in-process to finished goods on a daily basis. Strutt had the most sophisticated inventory control system, recording varieties of cotton in their raw material inventory at average and standard cost and tracking all materials through various stages of production.

Standard costing in the Industrial Revolution predated literature on the subject by the scientific management writers Church, Whitmore, and Emerson. Individual entrepreneurs conducted intricate experiments leading to the development of cost standards for various processes. Although the overall quality of the standards utilized did not compare with those post-1885 (Chatfield 1977), several firms (Boulton & Watt, Carron, Strutt) employing standard costs for making departmental comparisons anticipated Norton (1900), perhaps the nineteenth century's outstanding theoretical exponent of standard costing.

IV. Cost Management Motivations

The question arises as to the motivations behind these observed developments in cost management within firms of the Industrial Revolution. Raistrick (1970) documented a lengthy period of declining market prices that affected the iron industry firms from 1808 to 1830. By 1821, competition between Crawshay and other Welsh iron dealers led to depression in the Bristol market. The struggle to cope with falling market prices caused firms in the industry to undertake careful investigations of their ironmaking costs (Raistrick 1970). At various periods during the Industrial Revolution, intense competition squeezed available profit margins and led iron producers to scrutinize costs closely. Such reactions were evident at Carron, which developed cost management practices aimed at preserving both profitability and liquidity (Fleischman and Parker 1990).

A further motivation was the funding requirements for plant and equipment expansions. Funds had to be obtained from retained earnings, capital injections by existing partners, admission of new partners, or borrowing. As happened at Carron, the capital costs of extensions often exceeded initial projections and stretched the financial resources of the firm to the limit. Boulton & Watt, for example, engaged in building and plant expansions from 1800 to 1803 and employed cost management techniques to obtain greater economy and efficiency of operations. The firm introduced new production methods that closely combined technical and economic considerations. Roll (1968) argued that as the firm's monopoly ceased after 1800 and price competition became more intense, its costing and bookkeeping became more exact.

In his study of Wedgwood, McKendrick (1970) observed that major product-cost investigations were launched when, as in 1772, the market for the firm's products showed signs of slackening. Josiah Wedgwood's costing techniques helped him identify high margin product lines. Once having discovered the value of accurate costing in periods of depressed markets, he retained those techniques as part of his routine management processes.

The motivating effect of a competitive market is also evident in Rimmer's (1960) account of Marshall's languishing sales during the recession of 1827. Another fall in market prices led to negative margins from 1840. In fact, as a result of competition among Leeds spinners, a maturing Irish flax industry, and French tariffs, 17 spinning firms failed in the years 1840–1841, leaving the industry dominated by a few surviving firms. Once again, close attention to cost management practices was the order of the day.

V. Iron and Textile Industry Factors in Industrial Revolution Cost Management

Because of its orientation toward large manufacturing enterprises, this survey is dominated by the 13 textile and six iron business concerns. This concentration of economic activity affords the opportunity to investigate industry-specific factors in the development of cost-accounting. Both industries shared a number of common characteristics. They were largely comprised of family firms or partnerships whose main source of capital was ploughed-back profits (Pawson 1979). These firms were often headed by entrepreneurs who were innovators both in technology and cost-operations management. Both industries required large capital expenditures on plant and equipment and a sizeable work force in comparison with other industries (Lee 1986). Access to cheap supplies of raw materials and the employment of technology to reduce costs through efficient use of raw materials were central to profitability (Farnie 1979; Hyde 1977). Overseas trade and the development and exploitation of new markets were also essential (Chapman 1987; Harris 1988). Production activities tended to be geographically concentrated, which necessitated interrelationships with the transport industry (Lee 1986; Pawson 1979).

Certain similarities in methodology between textile and iron firms are evident in the foregoing discussion of cost management activity. Businesses in both industries utilized costing data for nonroutine decision making. Crowley (iron) and Oldknow (textile) made vertical integration decisions, and Darby (iron), Marshall (textile), and Greg (textile) evaluated technological innovations and improvements. Cost analyses were prepared to justify decisions to lease or buy steam engines, and statistics were maintained to evaluate the performance of multiple engines (by the iron industry firms Dowlais and Darby and by the Gott textile firm).

Regular cost comparisons were a feature of the accounting system in both industries. Marshall (textile) and Darby (iron) compiled annual comparative statistics on their multiple plants. Strutt (textile) made managers responsible for operations in the various carding and spinning rooms, and the Crawshay (iron) and Carron (iron) owners demanded weekly blast furnace reports for similar evaluations.

Enterprises in both industries reflected cost-accounting sophistication in daily operations. The Carron (iron), Darby (iron), and Greg (textile) archives contain frequent and intricate product costings. Overhead was applied to the product on the basis of a predetermined formula at Carron (iron), Dowlais (iron), Marshall (textile), and Strutt (textile). Production scheduling was done at Philips (textile) and Crowley (iron). Care was taken in process evaluation at Carron (iron) with product line decisions and blast furnace records, at Marshall (textile) with daily performance cards for all machines and workers, and at Strutt (textile) with precise expense accounts. Standard costing and inventory control techniques were commonplace.

Cost-accounting differences between textile and iron concerns appear to have been a function of firm complexity. Because there were more major technological innovations in iron, cost analyses tended to be more frequent and lengthier. There were also more lease-or-buy and transportation subcontracting decisions. Notwithstanding, textile firms did experiment with new technology. Gott analyzed the cost benefit of mechanical rather than manual methods for converting raw wool into yarn and anticipated savings from cheaper spinning and finishing methods.

Firm size does not appear to have been correlated with the volume of cost management activity, though more work on smaller firms may show significant differences. Ashworth and Philips were smaller than McConnel & Kennedy, James Finlay, and Horrocks, but manifested superior cost management. McConnel & Kennedy, the huge Lancashire textile firm, and Newton, Chambers, the premier Sheffield iron concern, both exhibited a dearth of cost-accounting.

An obvious role for cost-accounting in both industries was the control of raw material costs. Major iron producers took pains to maintain cost and yield data on their mineral sources. Carron Company's accountant filed numerous reports during the 1770s about the firm's coal, limestone, and ironstone holdings and leases. Minerals were charged to user departments at transfer prices that reflected the expenses incurred in mining ore (Fleischman and Parker 1990). By the 1840s, Darby and Dowlais knew the mining costs and yields of the several varieties of coal and ironstone the firm utilized.

Textile firms were particularly attentive to raw materials and waste management. Benjamin Gott calculated cloth output from various wool mixtures. Henry Ashworth observed in his 1851 diary that using more Sea Island cotton would save a penny per pound on higher quality twist and would make the product line profitable. He also determined a waste standard of 3.15 ounces for each pound of yarn produced. The data categories maintained by Marshall and Strutt likewise included waste comparisons.

The foregoing comparative analysis has not revealed a significant number of differences between the costing procedures employed by textile and iron industry concerns. Firms analyzed costs for making routine and special decisions, engaged in cost comparisons, employed product-costing techniques (including overhead allocation), and attempted to control raw material costs. This outcome might have resulted from the common, periodic exposure of firms in both industries to strong competition, market-driven pricing, and narrow profit margins.

VI. Summary and Conclusions

Table 1 is a qualitative summary of the activities of the 25 firms in the cost management areas identified. We have indicated with the letter "A" those categories in which the surviving business records or secondary source material suggest meaningful activity undertaken. There was substantial activity in all the identified areas, with particular emphasis on cost control (72 percent) and standards utilization (80 percent). These techniques took many forms. At times they appeared rudimentary, but, in a number of instances, their development and application were quite sophisticated. What must be emphasized is our finding that these cost management methods were not simply historical financial accounting figures for post-event stewardship reporting to absentee owners. They reflected management's ongoing pursuit of cost and quality control, as well as profitability maintenance or improvement. The letter "E" in table 1 represents our further qualitative assessment of firms that reflected excellence or sophistication in terms of the breadth of cost management practices, the quality level achieved, the detail apparent in existing accounting records, and the number of references to cost issues over time. Carron Company and Boulton & Watt achieved an excellence rating in all four categories, closely followed by Dowlais, Marshall, and Strutt.

One problem in evaluating the quality of Industrial Revolution accounting lies in defining what constitutes good cost management. Although it is arguable that costing conventions or a general acceptance of costing techniques may be indications of sophistication, common usage is not an imperative. With the evolution of cost-accounting specialists still a century in the future and in the absence of textbook guidance, it was the entrepreneurs themselves, sometimes motivated by competitive markets and narrow profit margins, who appear to have developed techniques appropriate for their own enterprises. Standardization in this environment was therefore most unlikely.

We would also caution that costing methods do not have to conform to today's conventional wisdom to be considered acceptable. Provocative writings like Johnson and Kaplan's *Relevance Lost* (1987) raise questions about the capability of modern methods serving as yardsticks for judging the cost-accounting efficiency of older societies (Edwards 1989).

Costing records of the British Industrial Revolution await further investigation by accounting historians. Previous views of this period of cost management have not benefited from extensive consultation of such archival sources. Still, it must be remembered that Pollard (1965) claimed these accounting systems were incapable of providing reliable information regarding return on capital investment and evaluation of new technology and other innovations. We would concur to the extent that virtually all efforts at financial accounting were inadequate. Profit calculations on operations were suspect; there was "confusion between capital and revenue;" depreciation was not taken on assets though interest on owners' investment was treated as a cost; and cost data were rarely integrated into the financial records (Pollard 1965, 233). However, we do not find this indictment fatal. In writing about the early American experience, Johnson and Kaplan (1987, 9) observed that "the management accounting measures were designed to motivate and evaluate the efficiency of internal processes, not to measure the overall 'profit' of the enterprise." The evidence on early British industrial concerns likewise suggests an independent development of the managerial and financial accounting systems.

We disagree, however, with Pollard's assessment that Industrial Revolution costing universally failed to provide good data for decision making on technology options and other innovations. Pollard (1965, 265) conceded that the number of existing cost estimates was impressive, but doubted their accuracy and their *post-factum* utilization. He attributed a "cavalier attitude" about exact costing to large profit margins and the absence of competition. In this study, the matching of *ex ante* cost estimates and *post-factum* implementation of action has proved extremely difficult, given the gaps in most archival holdings. However, we have seen sufficient linkages to be more optimistic about their quantity and quality than Pollard. Ultimately, the difference in perception may be a function of sample. Pollard examined a wider variety of firm sizes and industries since he was investigating Industrial Revolution management generally. Searching out cost-accounting specifically, we focused on textiles and iron, highly competitive industries with low profit margins. Further research is required on smaller firms in less competitive environments.

Techniques introduced during the "costing renaissance" of the scientific management movement were superior to the cost-accounting of the Industrial Revolution, as indirect expenses became more accurately costed and as integration was achieved with financial reporting (Chatfield 1977; Garcke and Fells 1893). Of course, the orientation changed as cost data became increasingly utilized for improving and integrating manufacturing and marketing processes (Chandler 1959). However, we have seen practices in Industrial Revolution cost management previously thought not to have predated the 1880s, particularly in the areas of standard costing, managerial evaluation, and decision making. Much of the efficient cost management tradition the British entrepreneurs had inaugurated was lost by the 1880s due to the dearth of a literature, the decline of British managerial acumen in the mid-nineteenth century (Elbaum and Lazonick 1984), and the family capitalism motif so entrenched and limiting in U.K. industrial development (Chandler 1984). Subsequently, the accounting theorists of "the costing renaissance" independently developed and popularized methods that had fallen into disuse. The evidence indicates that the entrepreneurs of the British Industrial Revolution had good costing instincts for their environment, an awareness that advances our understanding of their significance in costing history.

Appendix
Manuscript Sources

Firm	*Location of Records*	*Types of Records*
Ashworth	Stafford PRO	Diaries, Expenditure/Production Reports, Mill Comparisons
Boulton & Watt	Birmingham Central	Engine Books, Estimates, Notebooks, Inventory, Letters
Cardwell, Birley	Rylands (Manchester)	Inventory, Expenditure/Production Reports
Carron	West Register House (Edinburgh) National Library (Edinburgh)	Cost Analyses, Production Reports, Minute Books Letters, Cost Estimates
Crawshay	National Library (Aberyswyth)	Letters, Memorandum Books
Crowley	British Museum Newcastle Central	Law Book Law Book administration

Appendix—*Continued*

Firm	*Location of Records*	*Types of Records*
Darby	Ironbridge Museum	Inventory, Minute Book, Valuations, P & L Statements
	Shrewsbury Public	Wastebook, Furnace Accounts, Inventory, Journals
	Shropshire PRO	Memoranda, Cost Analyses, Wage Books, Daily Accounts
Dowlais	Glamorgan PRO	Yields, Minute Books, Notebooks, Letters, Costings
Dugdale	Rylands (Manchester)	Job Costings, Expenses, Machinery Book
Finlay (James)	Glasgow University	Daybooks, Minutes, Ledgers, Letters
Gott	Leeds University	Notebooks, Engine Expenses, Expense Calculations
Greg	Manchester Central	Stoppages, Mill Ledgers, Stock Books, Cash Books
Herculaneum	Liverpool Central	Minute Book
Horrocks	Lancashire PRO	Trade Letters, Ledgers, Journals, Power Loom Notes
Marshall	Leeds University	Notebooks, Departmental Accounts
McConnel & Kennedy	Rylands (Manchester)	Output Accounts, Labor Accounts, Ledgers, Daybooks
Mona Mines	Bangor University	Notebooks, Smelting Expenses, Refining Charges
Nasmyth, Gaskell	Eccles Public	Daybooks, Sales, Letters
Newton, Chambers	Sheffield Central	Colliery Accounts, Letters, Production
Oldknow	Manchester Central	Output Reports, Costings, Mill Comparisons
Philips	Stafford PRO	Production Budgets, Process Costing
Robinson	Nottingham PRO	Cost Calculations, Mill and Operations Detail
Strutt	Manchester Central	Expense Books, Machine Ledgers, Mill-Specific Results
	Derby Central	Letters
Stubs	Manchester Central	Production, Employment, Inventory
Wedgwood	Keele University	Letters, Ledgers, Inventory, Pricings

References

Babbage, C. 1963. *On The Economy of Machinery and Manufactures*. New York: Augustus Kelley Publishers, reprint of 1835 edition.

Cantrell, J. A. 1984. *James Nasmyth and the Bridgewater Foundry*. Manchester: Manchester University Press.

Chandler, A. D. 1959. The beginnings of 'big business' in American industry. *Business History Review* 33 (Spring): 1–31.

——. 1984. The emergence of managerial capitalism. *Business History Review* (Winter): 474–503.

Chapman, S. D. 1987. *The Cotton Industry in the Industrial Revolution*. 2d ed. Studies in Economic and Social History. London: Macmillan Education.

Chatfield, M. 1977. *A History of Accounting Thought*. Huntington: Robert E. Krieger Publishing Co.

Church, A. H. 1916. *The Proper Distribution of Expense Burden*. New York: The Engineering Magazine Co.

Edwards, J. R. 1989. Industrial cost accounting developments in Britain to 1830: A review article. *Accounting and Business Research* 19 (Autumn): 305–17.

Edwards, R. S. 1937. Some notes on the early literature and development of cost accounting in Great Britain. *The Accountant* 97 (August-September): 193–95, 225–31, 253–55, 283–87.

Elbaum, E., and W. H. Lazonick. 1984. The decline of the British economy: An institutional perspective. *Journal of Economic History* (June): 567–83.

Elias, M., ed. 1960. *Iron in the Making*. London: Butler and Tanner, Ltd.

Euphemia, K., ed. 1973. *Letters of Josiah Wedgwood*. 3 vols. Didsbury: E. J. Morten, Ltd.

Farnie, D. A. 1979. *The English Cotton Industry and the World Market*. Oxford: Clarendon Press.

Fleischman, R. K., and L. D. Parker. 1990. Managerial accounting early in the British Industrial Revolution: The Carron Company, a case study. *Accounting and Business Research* 20 (Summer): 211–21.

Flinn, M. W., ed. 1957. *The Law Book of the Crowley Ironworks*. Durham: Andrews & Co.

——. 1962. *Men of Iron: The Crowleys in the Early Iron Industry*. Edinburgh: Edinburgh University Press.

Garcke, E., and J. M. Fells. 1893. *Factory Accounts, Their Principle and Practice*. 4th ed. London: Crosby, Lockwood & Son.

Harris, J. R. 1988. *The British Iron Industry, 1700–1850*. Studies in Economic and Social History. London: Macmillan Education.

Hudson, P. 1977. Some aspects of the 19th century accounting development in the West Riding textile industry. *Accounting History* 2 (Part 2): 4–22.

Hyde, C. K. 1977. *Technological Change and the British Iron Industry, 1700–1870*. Princeton: Princeton University Press.

Johnson, H. T. 1972. Early cost accounting for internal management control: Lyman Mills in the 1850s. *Business History Review* 46 (Winter): 466–74.

——, and R. S. Kaplan. 1987. *Relevance Lost: The Rise and Fall of Management Accounting*. Boston: Harvard Business School Press.

Lee, C. H. 1986. *The British Economy Since 1700: A Macro-Economic Perspective*. Cambridge: Cambridge University Press.

McKendrick, N. 1970. Josiah Wedgwood and cost accounting in the Industrial Revolution. *Economic History Review* 23, 2d series (April): 45–67.

Metcalfe, H. 1885. *The Cost of Manufacturers*. New York: John Wiley & Sons.

Norton, G. P. 1900. *Textile Manufacturers' Bookkeeping*. London: Simpkin.

Parker, R. H. 1969. *Management Accounting: An Historical Perspective*. New York: Augustus Kelley Publishers.

——. 1986. *The Development of the Accountancy Profession in Britain to the Early Twentieth Century*. London: Academy of Accounting Historians.

Pawson, E. 1979. *The Early Industrial Revolution: Britain in the 18th Century*. London: Batsford Academic.

Pollard, S. 1965. *The Genesis of Modern Management*. Cambridge, MA: Harvard University Press.

Raistrick, A. 1970. *Dynasty of Iron Founders*. Newton Abbot: David and Charles.

Rimmer, W. G. 1960. *Marshalls of Leeds Flax-Spinners 1788–1886*. Cambridge: Cambridge University Press.

Roll, E. L. 1968. *An Early Experiment in Industrial Organization.* New York: Augustus Kelley Publishers, reprint of 1930 edition.

Solomons, D. 1952. *Studies in Costing.* London: Sweet & Maxwell, Ltd.

Stacey, N. A. H. 1954. *English Accountancy 1800–1954.* London: Gee & Co., Ltd.

Stone, W. E. 1973. An early English cotton mill cost accounting system: Charlton Mills, 1810–1889. *Accounting and Business Research* 4 (Winter): 71–78.

Tann, J., ed. 1981. *The Selected Papers of Boulton and Watt.* Vol. 1. London: Diploma Press.

Trinder, B. S. 1974. *The Darbys of Coalbrookdale.* Chichester: Phillimore.

Unwin, G. 1924. *Samuel Oldknow and the Arkwrights.* Manchester: Manchester University Press.

Urwick, L., and E. F. L. Brech. 1964. *The Making of Scientific Management.* London: Sir Isaac Pitman & Sons, Ltd.

Wells, M. C. 1978. *Accounting for Common Costs.* Urbana: Center for International Education and Research in Accounting.

Yamey, B. S. 1960. The development of company accounting conventions. *Three Banks Review* 47 (Spring): 22–37.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 376–388

Earnings Announcements and the Convergence (or Divergence) of Beliefs

Dale Morse
Cornell University
Jens Stephan
University of Cincinnati
Earl K. Stice
Rice University

SYNOPSIS: Research on analysts' earnings forecasts has produced two major results. First, security analysts provide more accurate forecasts than do time-series models (Brown et al. 1987) and, second, analysts' forecasts become more accurate and less dispersed as the forecast horizon decreases (Brown et al. 1985). This study examines the effect of an annual earnings announcement on the dispersion of analysts' one-year-ahead forecasts. It seems logical that forecasts should be less dispersed after the release of a value-relevant publicly observable signal. However, we find the opposite; i.e., that forecasts become more dispersed than would be expected in the absence of an earnings announcement.

Security analysts' forecasts have often served as a proxy for the unobservable market expectation of earnings. Similarly, the dispersion of analysts' forecasts may proxy for the diversity of investor beliefs about future earnings. A number of studies have suggested that diversity of beliefs is important in security pricing as well as being a determinant of trading volume.[1] Insight into the effects of earnings announcements on the hetero-

---

[1] For examples of the effects of diversity of beliefs on security pricing, see Miller (1977), Jarrow (1980), Figlewski (1982), and Mayshar (1983). The effects on trading volume can be found in Hakansson et al. (1982), Varian (1985), Karpoff (1986), Comiskey et al. (1987), Ou et al. (1988), and Holthausen and Verrecchia (1990).

The authors would like to acknowledge the assistance of workshop participants in the Cornell Accounting, Stanford Summer Accounting Workshop, SUNY Buffalo, and the Wharton School. Also four anonymous referees have been very helpful in guiding paper revisions. Any errors are the responsibility of the authors.

Submitted March 1988.
Revisions received August 1988, April 1989, and May, July, and October 1990.
Accepted October 1990.

geneity of investors' beliefs will improve our understanding of how information gets impounded into prices and how information alters investors' portfolio decisions.

The Bayesian belief revision model developed in this study suggests that the surprise content of the signal and the diversity of the perceived precision of the signal are important factors in determining whether the information event will cause a convergence or divergence of forecasts. Holthausen and Verrecchia (1990) reach similar conclusions when they examine the effect of information on consensus. In their model, a decrease in consensus (increase in diversity of beliefs) is possible if there is disagreement about the effect of the signal on the value of the firm. Bamber (1987) argues that the surprise content of the announcement and disagreement about the interpretation of the signal are related; i.e., "more surprising or informative announcements are likely to spawn a wide variety of interpretations. . . ." Empirical results are consistent with the insights provided by our model. There is a greater divergence of forecasts when the earnings announcement contains a bigger surprise,[2] where surprise is defined as the difference between reported earnings and analysts' predictions of those earnings.

An alternative explanation for the empirical results is nonsynchronous updating of forecasts by analysts. By partitioning the sample on the length of time between the announcement date and the next IBES report, we are better able to identify which IBES report (the first or second after the announcement) contains updated forecasts. This partitioning provides a more accurate measure of changes in the dispersion of forecasts due to an earnings announcement.

Key Words: *Analyst forecasts, Bayesian revision, Heterogeneous beliefs.*

Data Availability: *We use the IBES Summary History tape and will respond to requests for data subject to the approval of Lynch, Jones and Ryan.*

THE remainder of this article contains the following. Section I develops a Bayesian model that defines the characteristics of an information event that can affect changes in the dispersion of beliefs. Section II describes the sample selection procedures and the experimental design; it is followed by empirical results in section III. Finally, section IV provides a summary and conclusions.

I. Bayesian Analysis of Belief Revision Due to a Public Information Event

The model developed in this section describes the characteristics of an earnings announcement that may change the dispersion of analysts' forecasts of future earnings.

---

[2] We are unable to empirically separate the surprise content of the earnings announcement from the disagreement about its impact on the value of the firm. Therefore, we cannot perform a direct test on the effect of disagreement about the value of the signal on changes in the dispersion of forecasts.

The model is Bayesian[3] in the sense that analysts have forecasts of future earnings that are revised when they observe the current earnings figure.[4] A comparison of the cross-sectional distribution of forecasts of future earnings before and after an earnings announcement determines the conditions under which beliefs converge or diverge.

The sequence of events in this model is as follows:

t_{-1}: analysts have forecasts of future earnings to be released at t_2,
t_0: current earnings are announced,
t_1: analysts produce revised forecasts due to the current earnings release, and
t_2: date at which the forecasted earnings are released.

Earnings are assumed to be normally distributed. The ith analyst's t_{-1} forecast consists of an estimate of the mean m_i and variance s_i^2 of earnings to be released at t_2. Individual mean forecasts are assumed to be drawn from a normal distribution with mean $\bar{m}$ and variance s_m^2. Therefore, the heterogeneity of prior beliefs is defined by s_m^2 and is compared to the dispersion of mean forecasts after the release of current earnings to determine if forecasts have converged or diverged. The current earnings number released at t_0 is represented by Y, which is assumed to be drawn from a normal distribution with the same mean as the future earnings released at t_2.[5] All analysts observe the same signal Y; however, the perceived variance of the distribution from which Y is drawn varies across individuals and is represented by s_{yi}^2. Allowing s_{yi}^2 to vary across individuals implies that analysts may disagree about the perceived accuracy of the signal and, therefore, its relative importance in revising their forecasts. Each individual will therefore revise his or her forecast in the following manner:

$$r_i = p_i m_i + (1 - p_i) Y, \tag{1}$$

where:

r_i = mean of the revised forecast for individual i,
$p_i = g_i / (g_i + h_i)$,
$g_i = \frac{1}{s_i^2}$ = precision of prior beliefs of individual i, and
$h_i = \frac{1}{s_{yi}^2}$ = perceived precision of the signal by individual i.

In order to determine whether forecasts have converged or diverged as a result of the t_0 earnings announcement, we compare the variance of r_i, defined as s_r^2, with the variance of prior forecasts, s_m^2. Given that p_i and m_i are drawn independently from separate distributions,[6] we can characterize the distribution of r_i with the distributional characteristics of p_i and m_i. In the following analysis, r_i, p_i, and m_i in equation (1) are replaced with $\tilde{r}$, $\tilde{p}$, and $\tilde{m}$, where the random variables represent the underlying distributions.

---

[3] The Bayesian decision-making framework has the advantage of simplicity. However, as noted by Holthausen and Verrecchia (1988, 84), it imposes the assumption that analysts are risk-neutral.

[4] Easton and Zmijewski (1989) find that a single analyst (*Value Line*) uses current earnings to revise expectations of future earnings. These forecast revisions appear to involve a weighting of prior beliefs and new information in the earnings announcement.

[5] This assumption implies that the time-series of earnings follows a random walk. However, analysts may have information other than the past series of earnings.

[6] This may not be completely accurate in terms of analysts' forecasts because analysts with extreme forecasts may have prior beliefs with higher variances. Since this would happen in both tails of the distribution, the effect on the interaction between $\bar{m}$ and $\bar{p}$ is unclear.

$$\tilde{r}=\tilde{p}\tilde{m}+(1-\tilde{p})Y, \tag{2}$$

or

$$\tilde{r}=(\tilde{m}-Y)\tilde{p}+Y,$$

and:

$$s_r^2=\bar{p}^2 s_m^2+(\bar{m}-Y)^2 s_p^2+s_m^2 s_p^2. \tag{3}$$

Although all individuals observe Y and all mean forecasts move toward Y, this is not sufficient to cause forecasts to converge.[7] The conditions for converging and diverging beliefs are determined by comparing s_r^2 and s_m^2, which are the variances of posterior and prior beliefs. Rearranging equation (3) yields the following conditions for convergence and divergence:

$$\frac{s_r^2}{s_m^2}=(\bar{p}^2+s_p^2)+s_p^2\frac{(\bar{m}-Y)^2}{s_m^2}. \tag{4}$$

The left side of equation (4) is the ratio of the posterior forecast dispersion to prior forecast dispersion; if the ratio is greater (less) than one, then forecasts have become more (less) dispersed. The right side of equation (4) is a function of the "standardized" surprise, $(\bar{m}-Y)^2/s_m^2$, of the current earnings signal Y. "Standardized" is placed in quotes because the term does not necessarily have a variance of one. The right side of equation (4) is also a function of the average adjustment factor $\bar{p}$ and the variation in that adjustment factor s_p^2. The magnitude of s_p^2 is limited by the average adjustment factor, which is a maximum when $\tilde{p}$ is distributed binomially; i.e., the maximum $s_p^2=\bar{p}(1-\bar{p})$. The first term on the right side of equation (4) is therefore constrained to be less than or equal to one.

Necessary conditions for the divergence of beliefs are a non-zero "standardized" surprise in the current earnings signal and variation in the relative weighting of priors and new information across individuals. Divergence is more likely to occur with larger values of the "standardized" surprise, $(\bar{m}-Y)^2/s_m^2$, and greater variation in the adjustment factor, s_p^2. The magnitude of "standardized" surprise required for divergence can be demonstrated graphically given values for s_p^2 and $\bar{p}^2$. Easton and Zmijewski (1989, table 1) have estimated $(1-p_i)$ for *Value Line* forecasts using quarterly data. Their median estimate for $(1-p_i)$ is 0.326 for one-quarter-ahead forecasts and 0.137 for two-quarters-ahead forecasts. The p_i for one-year-ahead *Value Line* forecasts therefore appears to be in the neighborhood of 0.9. Using a value of $\bar{p}=0.9$, equation (4) is graphed in figure 1 using both the maximum value $s_p^2=0.09$ and one-half the maximum value $s_p^2=0.045$. With the maximum variance for s_p^2, a "standardized" surprise in excess of 1.11 would cause divergence. With one-half the maximum variance for s_p^2, a "standardized" surprise in excess of 3.22 would cause divergence.

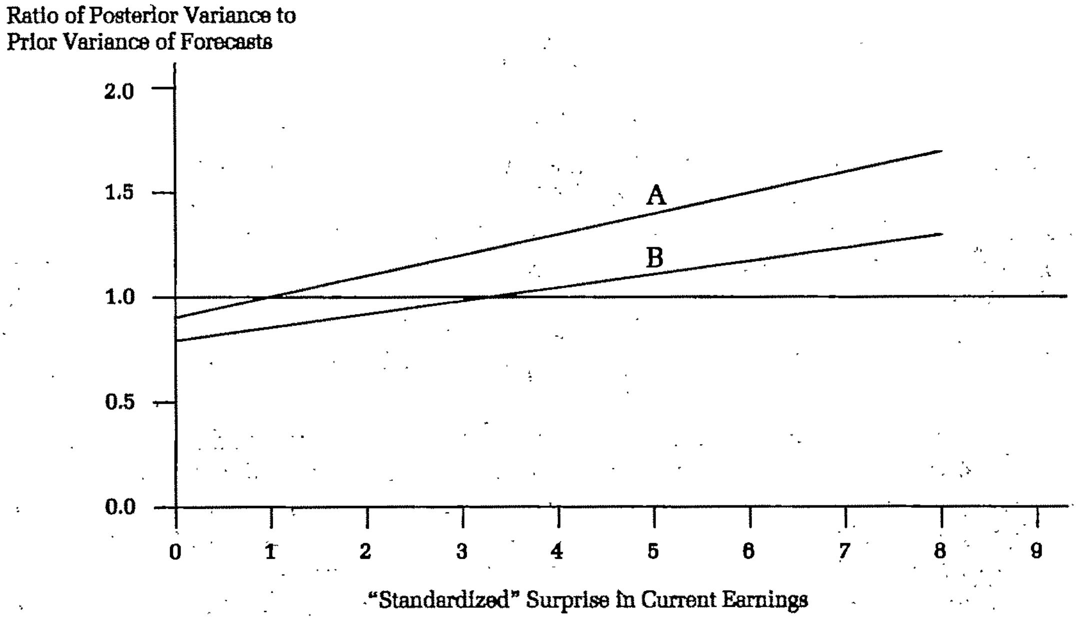
To determine the likelihood of observing diverging forecasts in our sample, we computed the "standardized" surprise of earnings for each firm year in the sample. "Standardized" surprise is operationalized in the model as the squared difference be-

---

[7] Given that every individual's mean forecast moves toward Y, it is somewhat surprising that forecasts can become more dispersed. The following example shows how. Let prior forecasts (at t_{-1}) be homogeneous, i.e., each analyst forecasts $\bar{m}$. A signal Y (not equal to $\bar{m}$) that is weighted differently by each analyst would result in revisions that are more dispersed than the prior forecasts.

Figure 1

Conditions for Convergence or Divergence of Forecasts
Assuming Average Adjustment Factor = 0.9



Ratios > 1.0 (< 1.0) imply a divergence (convergence) of forecasts
A: Assumes maximum variance of adjustment factor
B: Assumes one-half of the maximum variance of adjustment factor

tween the mean forecast of next year's earnings and the current year's earnings divided by the variance of next year's earnings forecasts. The cumulative frequency distribution for "standardized" surprise in our sample is:[8]

"Standardized" surprise:	0.00	1.00	2.00	3.00	4.00	10.00
Cumulative frequency:	0.00	0.66	0.76	0.81	0.83	0.90

Therefore, in our sample approximately 19 percent of the observations contain a "standardized" surprise sufficiently large (>3.22) to cause a divergence of forecasts (using one-half the maximum variance of the adjustment factor).

Equation (4) also implies that there is a positive linear relationship between the percentage change in forecast variances and the "standardized" surprise. This suggests that s_p^2 and $\bar{p}^2$ can be estimated through the coefficients of a regression.[9] In this study we

---

[8] We also examined the frequency of the "standardized" surprise defined as the square of the prior mean forecast of current earnings (instead of future earnings) less actual current earnings divided by the variance of current earnings forecasts. The cumulative frequency distribution for "standardized" surprise measured this way is:

"Standardized" surprise:	0.00	1.00	2.00	3.00	4.00	10.00
Cumulative frequency:	0.00	0.50	0.59	0.66	0.70	0.90

[9] A number of problems are apparent. First, $\bar{p}$ and s_p^2 are likely to vary across firms and across time. Second, both dependent and independent variables will be measured with error, as well as being highly skewed, which would require transformations to satisfy the statistical properties of regression analysis.

estimate equation (4) and test whether the coefficient on "standardized" surprise is positive.

II. Sample Selection and Experimental Design

The empirical work examines the impact of an annual earnings announcement on the dispersion of analysts' forecasts of future earnings. Brown et al. (1985) found a pattern of declining variances in analysts' forecasts as the forecast horizon grew shorter. We test whether this pattern is interrupted by an earnings announcement.[10] The IBES Summary History tape (provided by Lynch, Jones & Ryan) is used to obtain earnings and analysts' forecasts of earnings. Forecasts were collected for the three months prior to and five months after an annual earnings announcement during fiscal years ending forecasts in each of the eight months was required for inclusion in the sample. This limits the sample to primarily larger and more carefully watched firms.[11] COMPUSTAT was used to obtain earnings announcement dates. The joint requirements for obtaining forecasts and announcement dates yielded 5,752 firm-year observations.

The delay by some analysts in updating their forecast and the delay of its appearance in the IBES Summary History tape pose difficult problems in using the variance of analysts' forecasts. For example, if an announcement causes all analysts to increase their forecasts by two dollars, there would be no change in the variance of analysts' forecasts. However, if only one-half of the analysts update their forecast in response to a signal, there will be an apparent increase in the variance of forecasts in that month followed by a corresponding decrease in variance when the remaining analysts update their forecasts. The delay between a forecast update and its appearance in the IBES report represents a similar problem. O'Brien (1988) found an average delay of 34 trading days between an update and its inclusion in IBES.[12] Therefore, the full impact of updated forecasts as reported by IBES may occur one to two months following the announcement. Both delays will affect the month in which the update is reflected in IBES and can have an important effect on variance estimates, especially if there is a shift in the mean of analysts' forecasts.

Our solution to this problem is to partition the sample on the basis of the elapsed time from the earnings announcement to the next IBES report. As this period lengthens, the next IBES report is more likely to contain updated forecasts. As this period grows shorter, the second IBES report following the announcement is more likely to contain updated forecasts.

III. Empirical Results

Table 1 reports descriptive statistics for the variance of analysts' forecasts during the months surrounding an earnings announcement (month -1 ($+1$) is the IBES report immediately preceding (after) the announcement). The decline in variance reported by

---

[10] Ziebart (1990) finds that the decline in dispersion of forecasts is not evident in small samples and at the individual firm level. Our study examines the decline in dispersion of forecasts for large samples only.

[11] The median sales for firms in our sample were approximately 2.5 times greater than the median sales of other firms reported by COMPUSTAT.

[12] The standard deviation of the 34 trading day lag found by O'Brien is 44.5 days. This indicates that the distribution is highly skewed and that some forecasts have very long lags before being reported by IBES. The median lag is therefore much less than 34 days.

Table 1
Variance in Analysts' Forecasts[a] by Month

Variance of Analysts' Forecasts	*Month Relative to Earnings Announcement*[b]							
	−3	*−2*	*−1*	*+1*	*+2*	*+3*	*+4*	*+5*
First Quartile	0.0081	0.0081	0.0064	0.0064	0.0064	0.0064	0.0049	0.0049
Median	0.0289	0.0289	0.0256	0.0256	0.0225	0.0225	0.0196	0.0196
Third Quartile	0.1089	0.1024	0.0961	0.0900	0.0841	0.0784	0.0784	0.0729

Sample size = 5,752 firm years

[a] IBES reports the standard deviation of analysts' forecasts rounded to the nearest cent. Thus, for example, a forecast variance of 0.0064 means that IBES reported a forecast standard deviation of eight cents per share.

[b] Month −1 (+1) is the IBES report preceding (following) the announcement.

Brown et al. (1985) can be seen in table 1. This study focuses on the effect of an earnings announcement on this declining pattern. It is difficult, however, to draw any conclusion from table 1 because the cross-sectional comparisons are not standardized. For example, the first quartile in table 1 probably includes firms with small stock prices and earnings and has little to do with the dispersion of forecasts.

The data are standardized by first computing month-to-month percentage changes in the variance of analysts' forecasts for each firm year. These monthly percentage changes are then standardized to mean zero and standard deviation one. The data for each firm year, therefore, consist of seven changes in the variance of analysts' forecasts that sum to zero and have a standard deviation of one. This approach controls for a linear decline in the variance of analysts forecasts and weights each firm year equally when the data are aggregated.[13]

Table 2 describes the number of forecast revisions in the seven months surrounding the earnings announcement. The first part of the table examines the number of increases, decreases, and no changes in the sequential variances of analysts' forecasts for each firm year. The statistical tests compare the proportion of increases, decreases, and no changes with respective proportions estimated from all other months in the table.[14]

The "no change" category in the first part of table 2 reflects either no forecast revisions having been made by any of the analysts for that firm month or revisions small enough that rounding forecasts to the nearest penny causes the appearance of no

---

[13] Similar results were obtained when nonparametric statistics based on seven ranks for each firm were used. Another way to standardize the dispersion of forecasts is to use the coefficient of variation. We chose not to use this measure due to potential problems when earnings are near zero.

[14] The z-statistic is approximated by:

$$\frac{q_i - q_j}{\sqrt{\left[\bar{q}\left(1-\bar{q}\right)\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\right]}}$$

where q_i equals the proportion of the group in month i; q_j equals the proportion of the group in all months excluding month i; $\bar{q}$ equals the proportion of the group in all months; n_i equals the sample size for each month (5,752); and n_j equals the sample size for all other months (34,512).

Table 2
Changes in Variances of Analysts' Forecasts Surrounding an Annual Earnings Announcement

Change in Variance	Months Relative to Earnings Announcement[a]						
	−3 to −2	−2 to −1	−1 to +1	+1 to +2	+2 to +3	+3 to +4	+4 to +5
Decreases	2,507	2,391	2,442	2,577	2,510	2,489	2,523
Percentage	43.6%	41.6%	42.5%	44.8%	43.6%	43.3%	43.9%
Z-statistic	0.45	−2.88	−1.42	2.46	0.54	−0.07	0.91
No Change	1,580	1,784	1,572	1,368	1,536	1,421	1,486
Percentage	27.5%	31.0%	27.3%	23.8%	26.7%	24.7%	25.8%
Z-statistic	1.44	8.01	1.18	−5.39	0.02	−3.68	−1.59
Increases	1,665	1,577	1,738	1,807	1,706	1,842	1,743
Percentage	28.9%	27.4%	30.2%	31.4%	29.7%	32.0%	30.3%
Z-statistic	−1.88	−4.61	0.39	2.54	−0.60	3.62	0.55
Mean std. % Change in Variance of Analyst Forecasts	−0.009	−0.013	0.012	0.005	0.001	0.019	−0.015
t-statistic	−0.81	−1.19	1.02	0.42	0.07	1.45	−1.17

[a] Month −1 (+1) is the IBES report preceding (following) the announcement.

Z-statistics are based on a large sample binomial test comparing proportions of specific months with proportions of all other months.

t-statistics are for differences of the mean from zero. Standard deviations are estimated separately for each month and are approximately one due to the standardization procedure.

change in variance. The number of "no changes" appears to be unusually high in the month prior to the announcement and unusually low in the second month following the announcement. The unusually large number of variance "increases" and "decreases" in the second month following the announcement are consistent with hypothesized delays in the updating of forecasts and delays in their inclusion in the IBES report. The mean standardized change in the variance of analysts' forecasts is reported in the second part of table 2. None of these means are statistically different from zero.

The delayed forecast revisions implied by the first part of table 2 suggest that adjustments in defining the event month are necessary. Therefore, the sample is partitioned on the basis of the length of time between the earnings announcement and the next IBES report date. Presumably, the longer this lag, the higher the probability that forecast revisions made in response to the earnings announcement are reflected in the first IBES report following the announcement. The number of forecast revisions by month for each group is presented in table 3.[15] There are proportionately fewer forecast revisions from month −1 to +1 for groups with short lags than for groups with long lags.

---

[15] The uniform slow increase in the number of forecasts (one to two percent each month for each lag) appears to be unrelated to the earnings announcement. Also, the lag appears to be unrelated to the number of forecasts per firm; for example, the correlation between lag and number of forecasts per firm is −0.003 and the mean lag for firms with more (less) than ten forecasts is 15.15 (15.32) days.

Table 3
Number of Analyst Forecast Revisions[a] by Length of Reporting Lag and by Month

	Months Relative to Earnings Announcement[b]						
Lag	*−3 to −2*	*−2 to −1*	*−1 to +1*	*+1 to +2*	*+2 to +3*	*+3 to +4*	*+4 to +5*
0–5 Days							
# Forecasts	13,421	13,856	14,211	14,429	14,659	14,829	14,983
# Revisions	3,590	2,953	2,547	3,111	2,603	3,203	3,372
% Revisions	26.7	21.3	17.9	21.6	17.8	21.6	22.5
6–10 Days							
# Forecasts	12,483	12,916	13,226	13,359	13,620	13,702	13,918
# Revisions	3,004	2,592	2,039	2,883	2,454	2,946	2,917
% Revisions	24.1	20.1	15.4	21.6	18.0	21.5	21.0
11–15 Days							
# Forecasts	8,761	9,006	9,160	9,264	9,434	9,491	9,579
# Revisions	2,031	1,802	1,523	1,967	1,719	2,184	1,958
% Revisions	23.2	20.0	16.6	21.2	18.2	23.0	20.4
16–20 Days							
# Forecasts	9,135	9,382	9,531	9,644	9,832	9,899	10,032
# Revisions	1,988	1,740	1,700	1,909	1,832	2,231	1,878
% Revisions	21.8	18.5	17.8	19.8	18.6	22.5	18.7
21–25 Days							
# Forecasts	15,851	16,259	16,490	16,691	16,937	16,973	17,175
# Revisions	3,606	3,257	3,168	3,256	3,126	3,972	3,403
% Revisions	22.7	20.0	19.2	19.5	18.5	23.4	19.8
26+ Days							
# Forecasts	10,868	11,158	11,381	11,540	11,784	11,875	12,058
# Revisions	2,490	2,164	2,506	2,282	2,547	2,916	2,629
% Revisions	22.9	19.4	22.0	19.8	21.6	24.6	21.8
All							
# Forecasts	70,519	72,577	73,999	74,927	76,266	76,796	77,745
# Revisions	16,709	14,508	13,483	15,408	14,281	17,452	16,157
% Revisions	23.7	20.0	18.2	20.6	18.7	22.8	20.8

Sample size = 5,752 firm years

[a] Number of individual analyst forecasts that differ from the forecast made by the same analyst in the prior month.

[b] Month −1 (+1) is the IBES report preceding (following) the announcement.

Accordingly, the previous tests are repeated on two subsamples: firm-years in which the lag between announcement and the next IBES report is short (0–10 days, #obs = 797 firm-years) and when the lag is long (25–35 days, #obs = 2,039 firm-years). *Ex ante*, one might expect most forecast revisions to be reflected in the month −1 to +1 IBES report if the time lag between the earnings announcement and the next IBES report were sufficiently long and in month +1 to +2 if the time lag were short (assuming there were no other information events).

Table 4 presents the results when the sample is partitioned on the basis of short and long lags between announcement and next IBES report. For the long-lag group, it appears that an unusually low number of "no changes" occurs between month −1 and +1, and, by implication, an unusually large number of revisions (especially variance in-

Table 4
Changes in Variance of Analysts' Forecasts Based on Time Lag Between Earnings Announcement and Next IBES Report

	Long Lag (25–35 Days) *Sample Size=797*		*Short Lag (0–10 Days)* *Sample Size=2,039*	
	Month		*Month*	
Change in Variance	*−1 to +1*	*+1 to +2*	*−1 to +1*	*+1 to +2*
Decreases	361	364	837	851
Percentage	45.3%	45.7%	41.0%	41.7%
Z-statistic	1.12	1.33	−2.01	−1.40
No Change	166	206	644	508
Percentage	20.8%	25.8%	31.6%	24.9%
Z-statistic	−3.71	−0.53	4.87	−1.77
Increases	270	227	558	680
Percentage	33.9%	28.5%	27.4%	33.3%
Z-statistic	2.37	−0.93	−2.53	3.22
Mean std. % Change in Variance of Analysts Forecasts	0.029	−0.044	−0.013	0.056
t-statistic	0.86	−1.35	0.71	2.59

Month −1 (+1) is the IBES report preceding (following) the announcement.

Z-statistics are based on a large sample binomial test comparing proportions of specific months with proportions of all other months.

t-statistics are for differences of the mean from zero. Standard deviations are estimated separately for each month and are approximately one due to the standardization procedure.

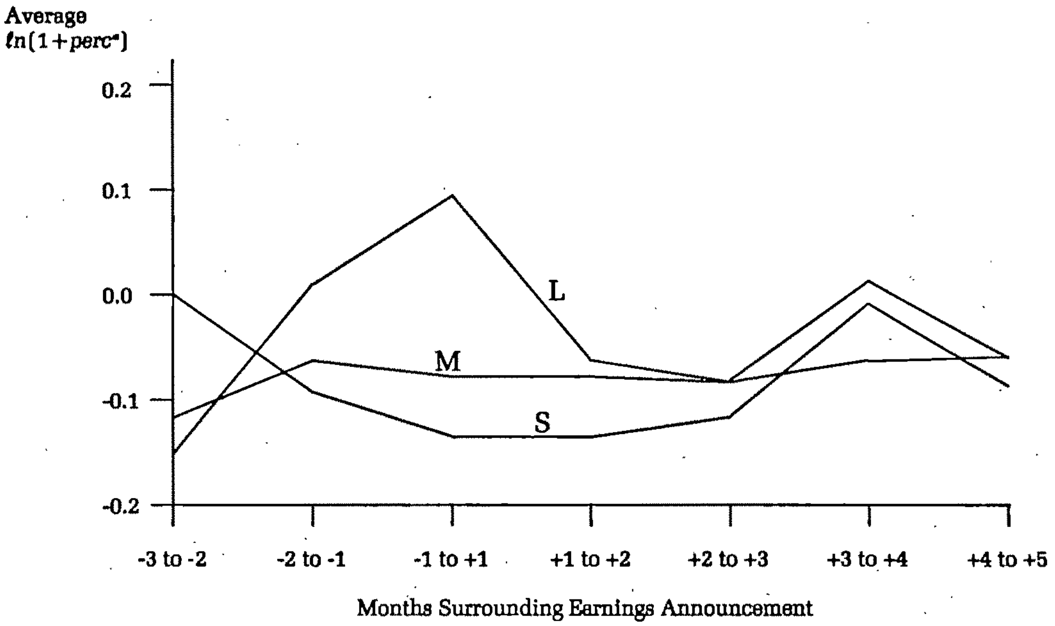
creasing revisions) is reported during that month. The short-lag group has fewer than expected revisions between month −1 and +1, which appears to indicate that this is not the month reflecting forecast revisions. Rather, the subsequent IBES report (month +1 to +2) has an unusually large number of variance increases and an unusually small number (marginally) of "no changes," which appears to indicate that this is the event month. To determine whether annual earnings announcements, on average, are associated with an increase or decrease in the dispersion of forecasts, we examine the mean standardized percentage change in the variance of analysts' forecasts (part 2, table 4). The only statistically significant change in variance of forecasts occurs in the predicted event month (+1 to +2) for the short-lag group. It is important to realize that the results in table 4 reflect deviations from the normal decline in the variance of analysts' forecasts. Decreases in forecast variances still outnumber increases in all months near the earnings announcement.

The Bayesian model of belief revision suggests that increases in forecast revisions are more likely to occur with larger "standardized" earnings surprises. Panels A and B of figure 2 illustrate the mean log of percentage changes in variance of forecasts for large, medium, and small "standardized" earnings surprises for long-lag and short-lag groups, respectively. The results are qualitatively similar to those in table 4. The long-lag (short-lag) sample exhibits a large increase in the mean percentage change in vari-

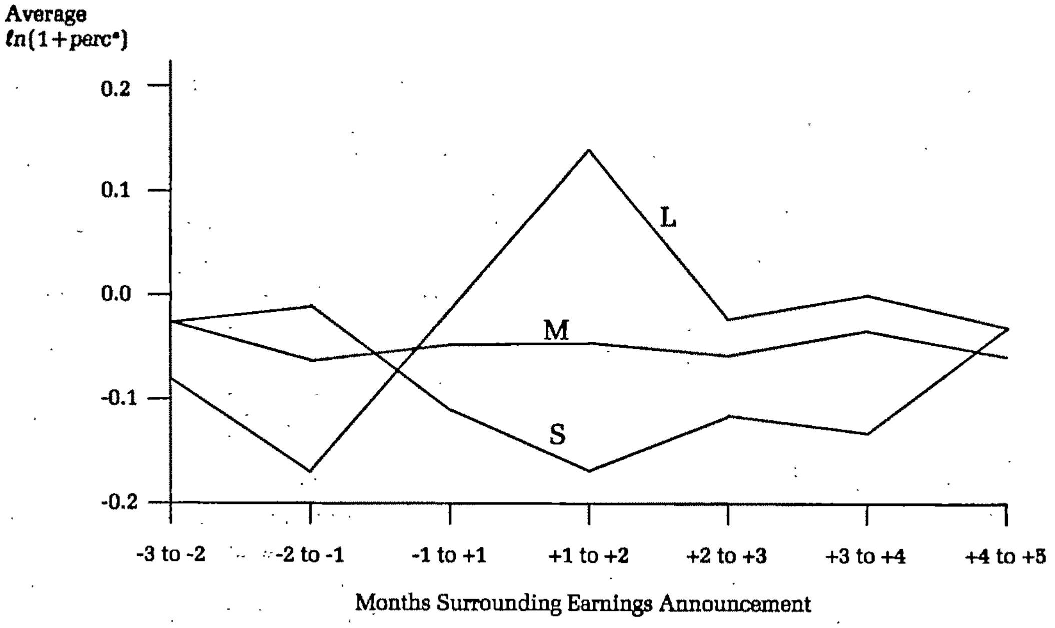
Figure 2

Mean Percentage Change in Variance of Forecasts Surrounding Earnings Announcements for Large, Medium, and Small Surprises

Panel A. Long-Lag Sample:



Panel B. Short-Lag Sample:



* perc=percentage change in variance of forecasts
Surprise: L=Large M=Medium S=Small

ance of forecasts in the first (second) IBES report after the earnings announcement. The difference between the mean percentage change in variance for the large "standardized" surprise group and the small "standardized" surprise group is statistically significant at the 0.05 level for month -1 to $+1$ ($+1$ to $+2$) for both the long-lag and short-lag samples.

We also examine the relationship between the percentage change in forecast variance and the "standardized" surprise with a cross-sectional regression in month -1 to $+2$ for the long-lag sample and month $+1$ to $+2$ for the short-lag sample. The regression is of the form:

$$\ell n(1+perc_{it})=b_0+b_1 \bullet \ell n(surp_{it})+e_{it},$$

where:

$perc_{it}$ = percentage change in variance of forecasts during event month for firm i in year t,
$surp_{it}$ = "standardized" earnings surprise for firm i in year t (refer to eq. [4]), and
b_0, b_1, e_{it} = estimated coefficients and residual.

The t-statistics for b_1 for the long-lag (short-lag) group are 3.65 and 8.34, respectively. Rank order correlations between the percentage change in variance during the event month and the "standardized" surprise were 0.168 for the long-lag group and 0.195 for the short-lag group. Both correlations are significant at the 0.001 level. This provides additional evidence on the association between the surprise content of the earnings announcement and change in dispersion of forecasts.

Even with partitioning the sample on the basis of time lag between the earnings announcement and the next IBES report, it is still possible that there exists more than one event month for each group, for example, one-half of the analysts' revisions may be included in the predicted event month IBES report and the other half may be included in the subsequent IBES report. The variance of forecasts would increase in the predicted event month but would then decline in the following month when the remaining analysts revise their forecasts. This effect would be most noticeable with large changes in mean forecasts, i.e., with large earnings surprises. Figure 2 indicates that the percentage *change* in forecast variance does not drop below the normal decline in variance in the month following the event month (month -1 to $+1$ for the long-lag group and month $+1$ to $+2$ for the short-lag group).[16]

IV. Summary and Conclusions

A Bayesian model is developed to determine conditions for the convergence or divergence of forecasts due to the release of a public signal. Divergence can occur only if there is a surprise in the public signal *and* the weighting on priors and the signal differs across analysts. Larger surprises are more likely to cause divergence. The empirical results of this study suggest that an earnings announcement disrupts the normal convergence of forecasts. Consistent with the model, we find a greater divergence of forecasts when the surprise in the earnings announcement is larger.

---

[16] Note that the Y-axis is defined as the *change* in the variance of forecasts. Therefore, a positive change in variance in the event month followed by a zero change in variance (as shown in fig. 2) implies a divergence of forecasts followed by no change in the dispersion of forecasts.

The model and empirical design offer some interesting directions for future research. The average degree to which analysts use an earnings report $\bar{p}$ and their disagreement on the importance of the earnings report s_p^2 are important variables in determining whether forecasts converge or diverge. These variables are likely to vary across firms and different types of annual reports. For example, accounting changes may influence the use of an annual report and therefore $\bar{p}$ and s_p^2. The existence of other information sources would also affect those variables, and their effect on the convergence or divergence of forecasts would provide additional insights.

References

Bamber, L. 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *The Accounting Review* 62 (July): 510–32.

Brown, P., G. Foster, and E. Noreen. 1985. *Security Analyst Multi-Year Earnings Forecasts and the Capital Market.* AAA Studies in Accounting, Research #21.

Brown, L., P. Griffin, R. Hagerman, and M. Zmijewski. 1987. An evaluation of alternative proxies for the market's assessment of unexpected earnings. *Journal of Accounting & Economics* 9 (July): 159–93.

Comiskey, E., R. Walkling, and M. Weeks. 1987. Dispersion of expectations and trading volume. *Journal of Business Finance & Accounting* (Summer): 229–39.

Easton, P., and M. Zmijewski. 1989. Cross-sectional variation in the stock market response to accounting earnings announcements. *Journal of Accounting & Economics* 11 (July): 117–41.

Figlewski, S. 1982. Information diversity and market behavior. *Journal of Finance* 37 (March): 87–102.

Hakansson, N., G. Kunkel, and J. Ohlson. 1982. Sufficient and necessary conditions for information to have social value in pure exchange. *Journal of Finance* 37 (December): 1169–81.

Holthausen, R., and R. Verrecchia. 1990. The effect of informedness and consensus on price and volume behavior. *The Accounting Review* 65 (January): 191–208.

Jarrow, R. 1980. Heterogeneous expectations, restrictions on short sales and equilibrium asset prices. *Journal of Finance* 35 (December): 1105–13.

Karpoff, J. 1986. A theory of trading volume. *Journal of Finance* 41 (December): 1060–88.

Mayshar, J. 1983. On divergence of opinion and imperfections in capital markets. *American Economic Review* (March): 114–28.

Miller, E. 1977. Prices, uncertainty and divergence of opinion. *Journal of Finance* 32 (September): 1151–68.

Morse, D. 1984. An econometric analysis of the choice of daily versus monthly returns in tests of information content. *Journal of Accounting Research* 22 (Autumn): 605–23.

O'Brien, P. 1988. Analysts' forecast as earnings expectations. *Journal of Accounting & Economics* 10 (January): 53–83.

Ou, J., J. Sepe, and N. Ushman. 1988. The relationship between the dispersion of analysts' forecasts and abnormal trading volume surrounding annual earnings announcements. Working paper, Santa Clara University.

Stickel, S. 1989. The timing of and incentives for annual earnings forecasts near interim earnings announcements. *Journal of Accounting & Economics* 11 (July): 275–92.

Varian, H. 1985. Divergence of opinion in complete markets: A note. *Journal of Finance* 40 (March): 309–18.

Ziebart, D. 1987. Evidence regarding divergence of analysts' forecasts of annual earnings per share: Does consensus increase as the forecast horizon declines? Working paper, University of Illinois.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 389–401

Volume of Trading and the Dispersion in Financial Analysts' Earnings Forecasts

Bipin B. Ajinkya
University of Florida
Rowland K. Atiase
University of Texas at Austin
Michael J. Gift
University of South Florida

SYNOPSIS: Varian (1985) and Karpoff (1986) showed analytically that trading volume is positively related to the degree of differing beliefs. This study provides empirical evidence on the postulated relationship. The extent of disagreement or dispersion in financial analysts' forecasts of annual EPS for a firm is employed as the proxy for agents' differing beliefs about the firm's prospects. The revision in analysts' mean EPS forecasts, from one month to the next, is used to control for the volume effects of the net information signals emanating during the period.

While some researchers have addressed the relation between the level of trading and earnings announcements (Beaver 1968; Morse 1981), and between trading volume and the magnitude of earnings forecast errors (Bamber 1986, 1987), the impact of discordant expectations on trading volume has been the subject of more recent empirical examination (Comiskey et al. 1987; Ziebart 1990; Lang and Litzenberger 1989).

The hypothesis tested in this study posits that the fraction of outstanding common shares traded is a positive function of both forecast dispersion

We wish to thank Florence Atiase, Linda Bamber, Richard Dietrich, Alex Dontoh, Robert Freeman, Prem Jain, William Kinney, Jens Stephan, Steven Kachelmeier, James Ohlson, Katherine Schipper, Senyo Tse, and three anonymous referees for comments on earlier drafts. We also thank seminar participants at the 1987 A.A.A. Annual Meeting, Emory, Florida, Indiana, Georgia Tech, Michigan State, N.Y.U., and South Florida. The first author gratefully acknowledges the financial support provided by the Accounting Research Center, Fisher School of Accounting, University of Florida. The second author would like to acknowledge financial support provided by the Department of Accounting, Graduate School of Business, University of Texas at Austin, and the Ernst and Young Foundation. We commend Lynch, Jones & Ryan who provided access to the I/B/E/S data base.

Submitted September 1989.
Revisions received May 1990 and October 1990.
Accepted November 1990.

and mean forecast revision. While most previous studies of volume have concentrated on specific accounting disclosures, this study examines the trading associated with an almost continuous flow of information about the sampled firms that analysts implicitly use in their periodic revisions of earnings forecasts. Hence, tests of the model become possible at several times during the year, independent of formal accounting events/disclosures.

Monthly observations in each of the four years 1978 to 1981 are used for the sample of 420 calendar-year firms; total number of observations is 16,747 over 48 months. Generalized least squares (GLS) estimation is applied to observations pooled over time and cross-sections and for each of the four years separately. Ordinary least squares (OLS) estimations are also performed and reported for comparative purposes and also to assess the stability of the models in monthly cross-sections. The results indicate a significant positive association between the dispersion in analysts' forecasts of annual EPS and the volume of trading. A relatively stable and positive association is found even after controlling for the volume effects of the magnitude of monthly revisions in the mean analysts' (annual) EPS forecast. The evidence corroborates the theoretical result that the degree of heterogeneity in beliefs is a determinant of the intensity of trading.

Key Words: *Heterogeneous beliefs, Volume of trading, Dispersion in analysts' forecasts, Analysts' forecast revisions.*

Data Availability: *A list of the sample firms is available from the authors upon request. The analysts' forecast data are proprietary and can be obtained from Institutional Brokers Estimate System (I/B/E/S) of Lynch, Jones & Ryan, Inc. The rest of the data is available from sources identified in the text.*

THE remainder of the paper is organized as follows. In section I, we review the relevant research motivating our expectations. We also provide a brief discussion of the differences of our study from others. Empirical measures of the variables, sample description, and statistical procedures are discussed in section II, while section III presents the results. Concluding comments are provided in section IV.

I. Prior Studies and Operational Hypothesis

"There is no horse race without differing beliefs" (Pfleiderer 1984) is the central notion guiding this study.[1] Explicit theoretical support for a positive relationship between volume of trading and the degree of heterogeneity in investors' beliefs is provided in Varian (1985) in a simplified pure exchange setting. Similarly, Karpoff (1986) concludes that trading volume has essentially the same two causes—dispersion in prior expectations and idiosyncratic interpretations of information events.[2] Since it is difficult to construct a reasonable empirical proxy for idiosyncratic interpretations of information events, we concentrate on examining the relationship between trading volume and the degree of heterogeneity in investors' beliefs about the prospects for a particular firm.

---

[1] Early efforts (e.g., Lintner 1969) assumed that beliefs were exogenous, i.e., independent of the information in prices; and to the extent that agents' beliefs differed, trading is a natural utility-maximizing consequence: the optimists traded with the pessimists. With the advent of the concept of rational expectations (Grossman 1976, 1978), beliefs are assumed to be determined endogenously in the market. In the limit, prices can become so informative that all agents' beliefs converge. The apparent paradox here is that while agents presumably communicate their own knowledge through their offers to trade, the result (equilibrium) is such that no trading occurs (given optimal endowments, etc.). The reaction to this result has been to characterize a

Of the few earlier empirical studies, Comiskey et al. (1987) find a significantly positive relation between trading volume for the entire calendar year and the dispersion in analysts' forecasts of annual earnings measured at one point (March) in the year. In this study, we examine the relation between volume and dispersion with finer and more contemporaneous monthly observations. Furthermore, we provide evidence on the volume effects of the revision in mean analysts' earnings forecasts, as well as use this latter variable as a control in examining the relation between volume and dispersion.

Ziebart (1990) examines the relation between the change in volume and the change in dispersion associated with information from formal earnings announcements, while we examine the trading associated with an almost *continuous* flow of information about sampled firms that analysts implicitly use in their periodic revisions of annual EPS forecasts. Thus, Ziebart's (1990) study can be thought of as a special case of ours. Our tests directly derive from the theoretical results of Varian (1985). Furthermore, we address the stability of the hypothesized relations, by examining the related coefficients in four yearly cross-sections and 48 monthly cross-sections. (Lang and Litzenberger 1989 is very similar to Ziebart 1990). Thus, our study differs from other related studies in orientation and scope.

Operational hypotheses are stated next. Because data on the degree of differing beliefs about a firm's future stock returns are unavailable, we use the dispersion in financial analysts' forecasts of annual EPS as a surrogate. Underlying this choice is the assumption that periodic earnings are a sufficient statistic for the purpose of firm valuation (see Miller and Modigliani 1961; Ohlson 1983; Miller and Rock 1985).

In examining the relationship between the volume of trading and the dispersion in

---

"noisy" rational expectations equilibrium (see Grossman and Stiglitz 1980 and Diamond and Verrecchia 1981, among others), where the noise prevents the price from being a sufficient statistic for investors' private information; thus beliefs are allowed (assumed) to differ even in equilibrium—and trading ensues. Thus, whether beliefs are exogenous or endogenous, or a combination thereof, trading is rationalized as a consequence of differing beliefs.

[2] Hakansson et al. (1982, 1984) derived, as a by-product of their analysis of the social value of information in pure exchange, the following necessary and sufficient conditions for "no trading" to occur in a two-period world:

(i) investors' endowed portfolios provide a fully allocationally efficient allocation under the presignal beliefs,
(ii) investors' prior beliefs are essentially homogeneous,
(iii) investors' information structures (i.e., "interpretations" of signals) are essentially homogenous, and
(iv) utilities are time additive.

We share the authors' own characterization that these conditions are quite stringent and unlikely to be met in the real world. However, the conditions also provide a convenient framework for motivating the operational variables used in our study. As a practical matter, the converse of the "no trading" result would suggest that the degree to which the above identified conditions *do not* obtain in the typical securities market would lead to a proportionally greater volume of trading.

analysts' forecasts, prior empirical evidence suggests that it may be necessary to control for the volume effects of the monthly revision in mean analysts' EPS forecast. Prior studies (Beaver 1968; Morse 1981; Bamber 1986, 1987) find that unexpected earnings information (i.e., earnings forecast error) revealed by an earnings announcement is positively associated with trading volume. In the context of our study, we expect that the revisions in mean EPS forecasts to be positively related to the volume of trading.[3] We therefore employ the magnitude of the monthly revision in mean EPS forecast as a control variable.

In sum, the operational hypotheses posit that trading volume (measured as the fraction of a sample firm's outstanding shares traded during a month) is (1) a *positive* function of the dispersion in financial analysts' forecasts of annual EPS made during the current month, and (2) a *positive* function of the absolute revision in the mean of financial analysts' EPS forecasts from the previous month to the current month.[4]

II. Research Method

Data

The *Media General Daily Price History Tape* was used to obtain data on the volume of shares traded. The *Institutional Brokers Estimate System (I/B/E/S) History Tape* was used to acquire information on the dispersion of analysts' EPS forecasts and the month-to-month changes in the mean level of those forecasts. In addition, the *COMPUSTAT Annual Industrial Tape* was used to meet the first two sample selection criteria described below.

Sample firms met the following three criteria: (1) 31 December fiscal year end; (2) no stock splits or stock dividends over the period under investigation, that is, for the years 1978 through 1981; and (3) at least three forecasters contributing to the mean annual EPS forecast in any study month. The first criterion ensures that the forecast horizon is the same for all firms. The second criterion eliminated the potentially confounding volume effects of stock splits and stock dividends. The third criterion was employed to enhance the statistical stability of both the mean and the dispersion measure.[5] The period under investigation was restricted to 1978–1981 due to data limitations. The data set that met all selection criteria consisted of 420 firms with a minimum of 208 observations in February 1978, a maximum of 399 observations in December 1981, and a grand total of 16,747 observations.

---

[3] A reason for a positive association between the volume of trading and the earnings forecast error variable (mean revision in analysts' EPS forecast in this study) suggested by Beaver (1968) and Bamber (1987) is that the larger the unexpected earnings, the greater the scope for differential interpretation, and, hence, the greater the volume of trading. In our study, the information signals are not formal announcements, but rather information emanating during a period (month) that causes a revision in the mean of analysts' EPS forecasts.

[4] In an earlier version of the paper presented at the 1987 A.A.A. annual meeting, we specified the models with firm size as an additional independent variable. The results of that version are qualitatively similar to those reported in the current paper with respect to the coefficients of *DISP* and *MFREV* variables. Thus, the results are not affected by firm size. Upon serious reflection, we now deliberately specify the models without firm size because the size effect is redundant when we incorporate the two independent variables *DISP* and *MFREV*. Each of these two variables is highly correlated with firm size and therefore captures the size effect. Indeed, our expectation is borne out by the fact that the coefficient of the size variable (*VALUE*) in Ziebart (1990) is not significant.

[5] Replication of the analysis with minimum forecaster requirements of five and ten, although reducing sample size, resulted in substantially similar outcomes.

Measurement of Variables

The two independent variables used in this study are the dispersion in analysts' EPS forecasts for a firm (DISP) and the monthly revision in the mean of analysts' forecasts (MFREV).

The I/B/E/S service receives updates of current fiscal year EPS forecasts for each firm from several financial analysts. While I/B/E/S makes available to investors the EPS forecasts and other information on a weekly basis, it also compiles a *monthly* summary on each firm that indicates the number of forecasts, the mean and median EPS forecasts made in the current month for the current fiscal year, and the standard deviation across forecasters (analysts) during the current month. DISP (in each month) is the standard deviation across analysts divided by the absolute value of the mean EPS forecast during the current month. The standardization renders DISP scale-free across firms for the cross-sectional analysis conducted in each month.[6] This variable is similarly defined by Comiskey et al. (1987) and Elliott and Philbrick (1990).

DISP may be confounded by the fact that not all analysts make and transmit their EPS updates to I/B/E/S on precisely the same day. As a result, differing degrees of time lags can exist between the EPS updates provided by different analysts (Brown et al. 1985). Also, O'Brien (1988) provides empirical evidence consistent with the existence of differential lag in analysts' EPS forecasts.

If *both* the timing of updates and the information sets on which updates are based differ across analysts, then differences in EPS estimates would be due in part to the differential information available to analysts at different times. Thus, DISP would reflect something more than contemporaneous disagreement among analysts. Although the extent of the potential confounding effect is not verifiable, any differential lag-induced bias in DISP only adds noise to the DISP measure and works against our hypothesis of the positive relation between DISP and trading volume.

The absolute value of the difference between the mean EPS forecast made during the current and the prior month, divided by the closing stock price (see below),[7] was employed as the measure of mean forecast revision (MFREV). This measure surrogates for the net information signal emanating during the (monthly) period and is used to control for the volume effects of new information.

The dependent variable, FVOL (fraction of outstanding shares traded) was also measured on a monthly basis compatible with MFREV. An I/B/E/S date indicator

---

[6] As an alternative, though it does not have as much theoretical appeal, we scaled *DISP* by stock price rather than by absolute value of mean EPS forecast. The results are qualitatively identical. The correlation between *LFVOL* and *LDISP* (scaled by price) is not much different from the first two rows of table 2. The alternative correlations are:

	1978	*1979*	*1980*	*1981*
Spearman Correlation	0.182	0.141	0.066	0.128
Pearson Correlation	0.192	0.139	0.065	0.121

(All these correlations are significant at the 0.001 level.) The Spearman (Pearson) correlation between *LDISP* (deflated by absolute mean EPS forecast) and *LDISP* (deflated by stock price) is 0.95 (0.92).

[7] Replication of the analysis with the opening stock price at the beginning of the respective fiscal years as an alternative deflator resulted in substantially similar results. Also, for a subsample that excluded negative mean earnings forecasts, similar results were obtained with *MFREV* defined as the absolute value of the difference between the mean EPS forecast made during a current and the prior month, deflated by the mean forecast during the prior month.

(typically around the middle of the month) specifies the last date on which some forecast information could have been added to the set for that month. Daily trading volume was accumulated for the calendar month surrounding this hypothetical forecast date as a benchmark. The half-month period prior to the "forecast date" was included since most of the forecast revisions made known to I/B/E/S by the financial analysts were presumably also available (as advice) to their own investor clients during that period. The half-month period subsequent to the forecast is expected to capture the persistence of volume reaction to the forecast revisions since prior empirical evidence (Morse 1981; Bamber 1987) suggests that volume reaction to earnings information persists up to seven trading days (i.e., approximately one-and-one-half weeks) after disclosure. Karpoff (1986) provides a theoretical rationale for this persistence.

Model Specification and Estimation

The relationship investigated in this study is of the form:

$$FVOL=f(DISP, MFREV).$$

Two of the variables (*FVOL* and *DISP*) are significantly autocorrelated and economywide events imply contemporaneous correlation in *FVOL* across firms; evidence on the occurrence of both is presented below.[8] Hence, the generalized least squares (GLS) procedure was used to accommodate and adjust for the existence of these correlations in residuals. Consider the following linear model:

$$Y_{it}=\sum_{k=1}^{M} X_{itk}\beta_k+u_{it}, \tag{1}$$

where Y is the dependent variable; X_k are the independent variables; i is the number of firms in the cross-section; t is the time in months; β_k are the coefficients; and u is the error in the model. Following Parks (1967), we model the error structure (u_{it}) as a first-order autoregressive process with contemporaneous correlation between cross-section units. Parks (1967) shows that the estimator resulting from this procedure is consistent and asymptotically normally distributed. This GLS procedure was applied separately to the volume model(s) in *each* of the four years 1978 to 1981. Within a given calendar year, the time-series consisted of monthly observations during the 12 months January through December. For comparative purposes, we also applied OLS estimation to the observations pooled across months in each of the four years.

We used logarithmic transformations of both dependent and independent variables to reduce departures of the residuals from normality.[9] For *MFREV*, which had zero values for approximately 20 percent of the observations, we used $\ell n(0.001 + |\text{raw variable}|)$.[10] The transformed variables are labeled *LFVOL*, *LDISP*, and *LMFREV*.

---

[8] The autocorrelation in *DISP* between months (t) and ($t-j$) is extremely high (in the excess of 0.80) and suggests that a pure time-series based test of the relation between volume and dispersion would not be very powerful since the independent variable *DISP* is devoid of much variation in its observations over time for a given firm. This difficulty can be overcome by conducting a cross-sectional test across firms so that a reasonable variability in *DISP* is assured. However, simply using cross-sectional observations does not neutralize the existence of the high autocorrelation in both the volume and the *DISP* series. A reasonable solution to these problems is to use a GLS estimation procedure that removes the autocorrelation in both the dependent and independent variable series prior to cross-sectional estimation, as was done in this study.

[9] Comparing the degree of departure from normality of the *residuals* when raw variables were used in model estimations to that when log-transformed variables were used, the following results in favor of log transformations were observed. For example, for the estimations on observations for 1978, the average skewness,

The two models estimated were:

Model I:
$$LFVOL_{it} = b_0 + b_1 LDISP_{it} + u_{it}, \tag{2}$$

Model II:
$$LFVOL_{it} = d_0 + d_1 LDISP_{it} + d_2 LMFREV_{it} + u'_{it}. \tag{3}$$

In the estimations, observations for i firms were pooled over 12 months ($t = 1, \ldots, 12$ for January to December) in each calendar year. On these pooled time-series cross-section data, both the GLS and OLS estimations were conducted for the two models in each of the four years 1978 to 1981. The GLS procedure reduced the number of observations for two reasons. First, only firms with complete monthly data within a year could be used. Second, the computer algorithm employed to perform the particular GLS estimation had a matrix size limitation of a maximum of 235 observations within a cross-section. Since the number of firms with complete data in the "full" sample (denoted "F") exceeded this limit in 1979, 1980, and 1981, we randomly deleted the excess firms to form a "reduced" sample (denoted "R") that satisfied the limitation of the GLS computer program. The maximum number of pooled observations in the case of the "reduced" sample is 235×12 months $= 2{,}820$ in a given year. Hence, for comparative purposes, OLS regressions were estimated for both the full as well as the reduced samples.

Diagnostic tests (based on Model II) show that the starting average autocorrelation coefficients between residuals in adjacent months are 0.57, 0.55, 0.59, and 0.49 in 1978, 1979, 1980, and 1981, respectively. The starting average cross-sectional correlation coefficients between residuals are 0.24, 0.12, 0.11, and 0.06 in 1978, 1979, 1980, and 1981, respectively. The majority of the individual autocorrelation coefficients and several of the individual cross-sectional coefficients are significant at the 0.05 level. As noted above, the GLS estimation adjusts for these correlations in the residuals and thus is better specified than the OLS estimation.

---

kurtosis, and the D-statistic of the Kolmogorov-Smirnov normality goodness-of-fit test were 6.50, 74.13, and 0.24, respectively, for the case of raw variables, diminishing to 0.48, 0.64, and 0.05, respectively, for log-transformed variables.

To assure the readers that the log transformation of the independent variable *DISP* is not critical to obtain statistical significance, the correlation between *LFVOL* and *raw DISP* is as follows:

	1978	*1979*	*1980*	*1981*
Spearman Correlation	0.200	0.171	0.148	0.210
Pearson Correlation	0.165	0.088	0.052	0.262

These correlations are all significant at the 0.001 level and the Pearson correlations are not much different from the second row of table 2.

[10] We retained the approximately 20 percent of the observations where *MFREV* had zero values because in the yearly GLS estimations the sample size (firm months) had to be balanced (equal) for each month during a study year. Thus, imposing a requirement of deleting monthly observations with zero *MFREV* values would result in deleting the entire associated firm-years, which would reduce the observations by much more than 20 percent. Anyway, to demonstrate that the 0.001 adjustment does not change our results qualitatively, we estimated the correlation between *LFVOL* and *LDISP* after deleting the observations with zero monthly *MFREV* values in each of the study years and found that the results are not much different from the first two rows of table 2.

Table 1
Overall Summary Statistics
Distributions of Raw Dependent and Independent Variables
(16,747 Observations: 1978–1981)[1]

Percentiles and Mean	*Fraction of Shares Traded*[2] (FVOL)	*Dispersion of Earnings Forecasts*[3] (DISP)	*Mean Abs. Forecast Revision*[4] (MFREV)
Maximum	1.362	28.500	0.605
0.95	0.118	0.289	0.017
0.90	0.081	0.172	0.010
0.75	0.046	0.090	0.004
Median	0.027	0.048	0.002
0.25	0.017	0.028	0.001
0.10	0.011	0.018	0.000
0.05	0.008	0.013	0.000
Minimum	0.001	0.001	0.000
Mean	0.041	0.115	0.004

[1] Based on an average 349 firms per month × 12 months × 4 years.

[2] Number of shares traded per month as fraction of total number of shares outstanding.

[3] Standard deviation of monthly forecasts of annual EPS across financial analysts, divided by the absolute value of the mean forecast during the month.

[4] Absolute value of the difference in mean forecast for the current and previous months, divided by the current closing stock price.

III. Results

Table 1 summarizes various percentile values of the distributions of the raw dependent and independent variables based on the "full" set of 16,747 observations (12 months × 4 years × 349 firms on average). The range for each variable is quite large, and each variable exhibits some degree of positive skewness (means exceed medians). The logarithmic transformations significantly reduced the skewness (see fn. 9).

The mean volume traded each month (*FVOL*) is about 4.1 percent of the shares outstanding (while the median is 2.7 percent). The mean value of *DISP*, the measure of standardized dispersion (coefficient of variation), is 0.115, with a median of 0.048. Standardized mean absolute forecast revision (*MFREV*) has a mean and a median of 0.004 and 0.002, respectively. Analogous descriptive statistics for the reduced sample (10,608 total observations over the four years) are similar to those for the full sample and are not reported.

Table 2 provides Spearman rank-order and Pearson correlation coefficients between the dependent variable (*LFVOL*) and each of the two independent variables (*LDISP* and *LMFREV*), and between the two independent variables themselves. As expected, *LDISP* and *LMFREV* are both positively and significantly correlated with the fraction of shares traded (at the 0.001 level). *LDISP* and *LMFREV* are also positively and significantly correlated, perhaps reflecting different but related determinants of trading volume.

Next, we discuss the results for the two volume models based on data pooled across months in each calendar year. The following estimations are reported for each model:

Table 2
Variable Intercorrelations by Year[1]

Variable Pair[2]	*1978*	*1979*	*1980*	*1981*
r(*LFVOL, LDISP*)	0.200* (0.216)*	0.171* (0.182)*	0.148* (0.149)*	0.210* (0.201)*
r(*LFVOL, LMFREV*)	0.175* (0.199)*	0.142* (0.171)*	0.091* (0.107)*	0.164* (0.144)*
r(*LDISP, LMFREV*)	0.467* (0.576)*	0.484* (0.550)*	0.486* (0.566)*	0.519* (0.605)*
[Sample Size]	[n=3,527]	[n=4,177]	[n=4,406]	[n=4,637]

[1] The table shows Spearman rank-order (Pearson) correlation coefficients between pairs of log-transformed variables. The sample size is based on pooled observations over 12 months of each year, for the full sample. Statistical significance levels are denoted by * (<0.001).

[2] The variables are logarithmic transformations of the raw variables *FVOL* (fraction of outstanding common shares traded), *DISP* (dispersion of EPS forecasts across analysts), and *MFREV* (monthly revision in the mean EPS forecasts).

(1) generalized least squares for the reduced sample, labeled GLS(R), (2) ordinary least squares for the reduced sample, labeled OLS(R), and (3) ordinary least squares for the full sample, labeled OLS(F).

Table 3 presents the results for both Models I and II. For Model I, the coefficients of *LDISP* are, as postulated, positive and statistically significant at the 0.001 level across all estimations in all four years (except for GLS(R) in 1980, where it is significant at the 0.05 level). While the *t*-values for the coefficients of *LDISP* in the GLS(R) estimations are generally lower than their counterparts in the OLS estimations, they are nevertheless significant at conventional levels. Thus, the results for Model I support the inference that trading volume (*LFVOL*) is a positive function of the dispersion in analysts' forecasts (*LDISP*). Even though the unconditional relation between *LFVOL* and *LDISP* is significantly positive, the (positive) correlation between *LFVOL* and *LMFREV* reported in table 2 suggests the need to control for the volume effects of *LMFREV*. Model II enables an examination of this issue.

The results for Model II are also reported in table 3. The regression coefficients of *LMFREV* are positive and statistically significant at conventional levels, except for the GLS(R) and OLS(R) estimations in 1980 and for the OLS(R) estimation in 1981, where they are not significant. These results are consistent with prior empirical evidence (Bamber 1986, 1987) that the magnitude of earnings forecast error is positively associated with trading volume. The coefficients of *LDISP* in Model II are still positive and statistically significant at the 0.001 level across all estimations in all four years, except for the GLS(R) and OLS(R) estimations in 1980, where they are significant at the 0.05 level. Thus, even after controlling for the volume effects of *LMREV*, the dispersion in financial analysts' forecast of EPS is significantly positively related to the volume of trading.[11]

---

[11] Since the maximum values of the raw variables reported in table 1 are large relative to the 95th percentile values, we conducted an analysis of potential outliers. We replicated the (OLS) analysis after deleting observations associated with values of *FVOL*, *DISP*, and *MFREV* greater than 0.10, 1.0, 0.10, respectively. This resulted

Table 3

Regression Results for Models I and II[1]

(Model I: $LFVOL_{it} = b_0 + b_1 LDISP_{it} + u_{it}$)

(Model II: $LFVOL_{it} = d_0 + d_1 LDISP_{it} + d_2 LMFREV_{it} + u'_{it}$)[2]

		Model I		*Model II*		
Year	*Estimation* *[Sample Size]*	LDISP b_1	*Adj.* R^2	LDISP d_1	LMFREV d_2	*Adj.* R^2
1978	GLS (R) [n=2,148]	0.141 (14.96)**	N.A.	0.128 (14.48)**	0.127 (4.66)**	N.A.
	OLS (R) [n=2,148]	0.195 (10.81)**	0.051	0.122 (5.63)**	0.383 (6.01)**	0.066
	OLS (F) [n=3,527]	0.194 (13.11)**	0.046	0.134 (8.00)**	0.333 (6.10)**	0.055
1979	GLS (R) [n=2,820]	0.081 (6.85)**	N.A.	0.065 (3.69)**	0.078 (3.42)**	N.A.
	OLS (R) [n=2,820]	0.168 (10.03)**	0.034	0.127 (6.37)**	0.170 (3.80)**	0.038
	OLS (F) [n=4,177]	0.163 (11.99)**	0.032	0.113 (6.96)**	0.222 (5.60)**	0.040
1980	GLS (R) [n=2,820]	0.046 (2.78)*	N.A.	0.041 (2.35)*	0.032 (1.08)	N.A.
	OLS (R) [n=2,820]	0.058 (3.99)**	0.010	0.043 (2.48)*	0.065 (1.55)	0.010
	OLS (F) [n=4,406]	0.115 (10.00)**	0.022	0.100 (7.19)**	0.067 (1.87)*	0.022
1981	GLS (R) [n=2,820]	0.112 (4.73)**	N.A.	0.105 (4.54)**	0.057 (1.77)*	N.A.
	OLS (R) [n=2,820]	0.154 (10.00)**	0.033	0.138 (7.02)**	0.066 (1.30)	0.034
	OLS (F) [n=4,637]	0.160 (13.99)**	0.040	0.143 (9.97)**	0.072 (1.95)*	0.041

[1] The table shows estimated GLS and OLS coefficients and *t*-values (in parentheses) for the respective independent variables in the model. The qualifier (F) signifies the full sample, while (R) connotes a reduced sample that was necessitated by the size limitation of the GLS computer algorithm. Significance levels are denoted by * (<0.05) and ** (<0.001) for one-tailed tests. "N.A." indicates that an overall goodness-of-fit measure was not available in the GLS algorithm used.

[2] The variables are logarithmic transformations (see text) of the raw variables *FVOL* (fraction of outstanding common shares traded), *DISP* (dispersion of EPS forecasts across analysts), and *MFREV* (monthly revision in mean EPS forecast). The estimations are based on observations on firm *i* in month *t* during each calendar year.

in the deletion of approximately eight percent of the original observations overall. The results of this subsample are qualitatively identical to those of the original sample (table 3). Thus, the results are not unduly influenced by observations associated with the maximum range of values. It should be noted that the log transformations do moderate the effect of large raw values anyway.

Table 4
Percentiles of Monthly Estimations for Models I and II[1]
(48 Estimations from January to December for Years 1978–1981)

	Model I		*Model II*		
Percentile	LDISP b_1	*Adj.* R^2	LDISP d_1	LMFREV d_2	*Adj.* R^2
0.25	0.118 (3.02)**	0.022	0.087 (1.81)*	0.065 (0.47)	0.026
0.50	0.124 (3.66)**	0.038	0.138 (2.35)**	0.154 (1.01)	0.046
0.75	0.234 (4.38)**	0.053	0.160 (3.25)**	0.325 (1.76)*	0.062
mean [s.d.]	0.170 [0.053]		0.134 [0.060]	0.189 [0.198]	

[1] The table shows the distribution of estimated OLS coefficients and *t*-values (in parentheses) for the respective independent variables in Models I and II, for the full sample.

Model I: $LFVOL_i = b_0 + b_1 LDISP_i + u_i$
Model II: $LFVOL_i = d_0 + d_1 LDISP_i + d_2 LMFREV_i + u_i'$

The variables used are logarithmic transformations of the raw variables *FVOL* (fraction of outstanding common shares traded), *DISP* (dispersion of EPS forecasts across analysts), and *MFREV* (revision in mean analyst forecast). The percentiles of the 48 monthly estimations for Models I and II are determined by the *t*-values of the estimated coefficients of b_1 and d_1, respectively. Statistical significance levels are denoted by * (<0.05) and ** (<0.01). The table also reports the mean and standard deviation (s.d.) of the 48 estimated coefficients.

The regression results using data pooled over time-series and cross-sections (table 3) suggest that, while the magnitudes of the coefficients and related *t*-values are different for the GLS and OLS estimations, the inferences are qualitatively similar.[12] As a final step, in order to assess the stability of the relations in individual cross-sections, OLS regressions for the full sample were estimated for the two models separately in each month (a total of 48 estimations, 12 months × 4 years, per model). Table 4 presents the distribution of estimated coefficients and *t*-values (in parentheses) as well as the mean and standard deviation (s.d.) of the 48 monthly coefficients for the respective independent variables in Models I and II. The range of magnitudes of the estimated coefficients of the monthly regressions compare closely with those of the yearly regressions. The adjusted R^2s are slightly higher for the monthly regressions since the over-time (random) variance is less, compared to the results of table 3. The coefficient of *LDISP* is positive and statistically significant in Model I, at least at the 0.05 level, for all the 48 months. In the controlled version of the model, Model II, the significance of *LDISP* drops slightly to 39 months out of 48 (i.e., about 81 percent). Thus, while the various coefficients and their statistical significance levels vary somewhat over the different months and years, the inferences that can be drawn from the results are quite consistent from month to month.

---

[12] In addition to Models I and II, which are specified in the levels of the variables, we also ran the models specified in monthly first differences of consecutive months, as well as the models specified in seasonal first differences (by the month of year), but neither type of the first differences models yields systematic results. This is because in our data, the changes in the levels (i.e., whether first differences of consecutive months or seasonal first differences) are very small relative to the levels, so that there is effectively very little intertemporal variability to be explained (see also fn. 8).

IV. Concluding Comments

This study provides empirical evidence on the relationship between the volume of common stock trading and a proxy for the degree of dispersion in market agents' contemporaneous beliefs. By being sufficiently different in design and scope, our study can be thought of as a generalization of the other studies in this area.

In conclusion, the results support the following inferences:

1. The volume of trading in a firm's common stock is positively related to the dispersion in analysts' beliefs about the firm's prospects (as reflected in EPS forecasts).
2. Trading volume is positively related to the magnitude of the monthly revision in the mean of analysts' EPS forecasts. This result is consistent with prior empirical evidence.
3. The positive and significant relation between the volume of trading and the dispersion in analysts' EPS forecasts noted in (1) above holds even after controlling for the volume effects of the magnitude of the monthly revision in the mean analysts' EPS forecasts. A basic implication of this result is that future studies on trading volume need to consider the effects of not only the revision in the mean EPS forecasts (or, new information) but also the dispersion in the EPS forecasts (or, the heterogeneity in beliefs about the information variable).
4. The independent variables included in the analysis, while exhibiting conventional levels of statistical significance, do not explain a large proportion of the variation in trading volume across firms. The explanatory power of the model is nevertheless comparable to that achieved in typical market based event studies that use the stock returns metric. Finally, the validity of the inferences is subject to the *caveat* that dispersion in analysts' forecasts is a reasonable surrogate of its theoretical construct.

References

Bamber, L. 1986. The information content of annual earnings releases: A trading volume approach. *Journal of Accounting Research* 24 (Spring): 40–56.

———. 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *The Accounting Review* 62 (July): 510–32.

Beaver, W. 1968. The information content of annual earnings announcements. *Journal of Accounting Research* (Supplement): 67–92.

Brown, P., G. Foster, and E. Noreen. 1985. *Security Analyst Multi-Year Earnings Forecasts and the Capital Market. Studies in Accounting Research #21.* Sarasota, FL: American Accounting Association.

Comiskey, E. E., R. A. Walkling, and M. A. Weeks. 1987. Dispersion of expectations and trading volume. *Journal of Business Finance & Accounting* 14 (Summer): 229–39.

Cukierman, A., and D. Givoly. 1982. Heterogeneous earnings expectations and earnings uncertainty—Theory and evidence. Working paper, Tel Aviv University, Israel.

Diamond, D. W., and R. E. Verrecchia. 1981. Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics* 9 (September): 221–35.

Elliott, J., and D. Philbrick. 1990. Accounting changes and earnings predictability. *The Accounting Review* 65 (January): 157–74.

Givoly, D., and J. Lakonishok. 1979. The information content of financial analysts' forecasts of earnings. *Journal of Accounting & Economics* 1 (Winter): 165–85.

———. 1983. Divergence of earnings expectations: The effect on market response to earnings signals. Working paper, Tel Aviv University, Israel.

Grossman, S. 1976. On the efficiency of competitive stock markets where traders have diverse expectations. *The Journal of Finance* 31 (May): 573–85.

———. 1978. Further results on the informational efficiency of competitive stock markets. *Journal of Economic Theory* 18 (March): 81–101.

———, and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70 (June): 393–408.

Hakansson, N. H., G. J. Kunkel, and J. A. Ohlson. 1982. Sufficient and necessary conditions for information to have social value in pure exchange. *Journal of Finance* 37 (December): 1169–81.

———. 1984. A comment on Verrecchia's no trading 'theorem'. *Journal of Accounting Research* 22 (Autumn): 765–67.

Karpoff, J. M. 1986. A theory of trading volume. *The Journal of Finance* 41 (December): 1060–88.

Lang, L., and R. Litzenberger. 1989. Trading volume and changes in heterogeneous expectations. Working paper, New York University, NY.

Lintner, J. 1969. The aggregation of investor's diverse judgments and preferences in purely competitive security markets. *Journal of Financial and Quantitative Analysis* 15 (December): 347–99.

Miller, M., and F. Modigliani. 1961. Dividend policy, growth and the valuation of shares. *Journal of Business* 34 (October): 411–33.

———, and K. Rock. 1985. Dividend policy under asymmetric information. *The Journal of Finance* 40 (September): 1031–51.

Morse, D. 1981. Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research* 19 (Autumn): 374–83.

O'Brien, P. C. 1988. Analysts' forecasts as earnings expectations. *Journal of Accounting Economics* 10 (January): 53–83.

Ohlson, J. A. 1983. Price-earnings ratios and earnings capitalization under uncertainty. *Journal of Accounting Research* 21 (Spring): 141–54.

Parks, R. W. 1967. Efficient estimation of a system of regression equations when disturbance are both serially and contemporaneously correlated. *Journal of the American Statistical Association* 62 (June): 500–09.

Pfleiderer, P. 1984. The volume of trade and the variability of prices: A framework for analysis in noisy rational expectations equilibria. Working paper, Stanford University.

Varian, H. 1985. Differences of opinion and the volume of trade. Working paper, University of Michigan.

Ziebart, D. 1990. The association between consensus of beliefs and trading activity surrounding earnings announcements. *The Accounting Review* 65 (April): 477–88.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 402–416

Common Stock Returns Surrounding Earnings Forecast Revisions: More Puzzling Evidence

Scott E. Stickel
University of Pennsylvania

SYNOPSIS: The relation between changing expectations of earnings and changing security prices is a central issue in accounting and finance. In this article, I reexamine common stock returns surrounding earnings forecast revisions, using a large database of individual analyst forecasts, and provide new evidence on market expectations of revisions, on cross-sectional differences in price effects, and on the influence of confounding events.

In summary, my findings are that revisions affect prices, but prices do not immediately assimilate the information. Price reaction is greater when the percentage change in forecast is in the top or bottom five percent of the distribution of all forecast revisions. This price effect is not simply due to an association between revisions and earnings, dividend, or stock-split announcements. Surprisingly, prices continue to drift in the direction of the revision for about six months after the revision.

Another surprise is that price reaction does not incorporate some publicly available information. Stock returns immediately after individual analyst forecast revisions suggest that an analyst's current outstanding forecast is a better measure of the market expectations of the analyst's next forecast than an updated version (i.e., the analyst's current forecast updated for information revealed after the date of the current forecast but before the date of the next forecast).

I use this curious price reaction result to create an aggressive trading strategy that predicts changes in outstanding forecasts, in other words, a

---

I wish to express my gratitude to the following: Larry Brown, Bob Holthausen, Don Lewin, Craig MacKinlay, Jerry Zimmerman, and two anonymous referees for helpful comments; the KPMG Peat Marwick Foundation, Deloitte Haskins & Sells, and the Institute for Quantitative Research in Finance for financial support; and Zacks Investment Research, Inc., for supplying the analyst forecasts.

Submitted June 1989.
Revisions received March, July, and September 1990.
Accepted October 1990.

strategy that predicts price reactions. The difference in abnormal returns between securities predicted to perform best and worst is more than 13 percent every *six* months. Changes in beta do not explain these abnormal returns.

Key Words: *Earnings forecasts, Expectations, Information content, Trading strategy.*

Data Availability: *All data must be obtained from the sources identified in the text.*

THE relation between changing earnings expectations and changing security prices is a central issue in accounting and finance. In this article, I reexamine common stock returns surrounding earnings forecast revisions, using a large database of individual analyst forecasts, and provide new evidence on market expectations of revisions, on cross-sectional differences in price effects, and on the influence of confounding events.

Abdel-khalik and Ajinkya (1982) examine the effect of revisions on stock prices, using 288 mean forecast revisions by Merrill Lynch analysts in 1977–1978, and find significant abnormal returns during the publication week of the revisions, which is the week *after* the forecasts are released to employees and customers.[1] More recently, Lys and Sohn (1989), in a contemporaneous study, also examine the association between abnormal returns and revisions and find that revisions incorporate some, but not all, of the information in past prices and that revisions affect prices. My study offers a comprehensive analysis that uses 173,620 revisions for 1,465 firms by 1,869 analysts from 83 brokerage houses.

Subsequent stock price behavior warrants scrutiny because Givoly and Lakonishok (1979, 1980) and Hawkins et al. (1984) find abnormal returns long after large upward revisions. Givoly and Lakonishok, using the most active analyst reporting to Standard & Poor's *Earnings Forecaster* during 1967–1974, find abnormal returns of 2.7 percent during the *two months* following the revision month for 584 firms experiencing positive revisions of more than five percent of forecasted earnings. Hawkins et al. (1984) use a strategy that purchases the 20 stocks with the largest monthly increase in I/B/E/S mean consensus forecast and find abnormal returns of 14.2 percent in the *year* subsequent to revisions from 1975 to 1980. Brown et al. (1985) use about 20,000 I/B/E/S consensus revisions from 1976 to 1980 and find significant abnormal returns of −3.8 percent over the *11 months* following negative revisions, but no significant abnormal returns following positive revisions. Abdel-khalik and Ajinkya (1982) find no abnormal returns during the four weeks following the publication week, but use a relatively small sample of 288 revisions. In sum, evidence on price behavior after revisions is conflicting. Because of this and because of the magnitude of the returns reported by Givoly and Lakonishok (1979, 1980) and Hawkins et al. (1984), price behavior after revisions warrants reexamination with independent data, alternative methodologies, and different time frames.

---

[1] Other studies report abnormal returns during the month of the revision (e.g., Givoly and Lakonishok 1979, 1980). However, because analysts quickly respond to earnings announcements (see Stickel 1989), this evidence is confounded. Imhoff and Lobo (1984) use I/B/E/S monthly consensus revisions and eliminate revisions that are confounded by the release of other earnings information. However, the month 0 abnormal returns they report could be attributable to nonearnings information, released earlier within month 0, that triggers revisions included in the month 0 consensus revisions.

In summary, my findings are that revisions affect prices, but prices do not immediately assimilate the information. Price reaction is greater when the percentage change in forecast is in the top or bottom five percent of the distribution of all forecast revisions. This price effect is not simply due to an association between revisions and earnings, dividend, or stock-split announcements. Surprisingly, prices continue to drift in the direction of the revision for about six months after the revision.

Another surprise is that price reaction does not incorporate some publicly available information. Stock returns immediately after individual analyst forecast revisions suggest that an analyst's current outstanding forecast is a better measure of market expectations of the analyst's next forecast than is the analyst's current forecast *updated for* information revealed since the date of the current forecast but before the date of the next forecast.

I use this curious price reaction result to create an aggressive trading strategy that predicts changes in outstanding forecasts, in other words, a strategy that *predicts price reactions*. The difference in abnormal returns between securities predicted to perform best and worst is more than 13 percent every *six months*. Changes in beta do not explain these abnormal returns.

Although analysts have an effect on price that is independent of earnings announcements, my findings are analogous to studies that find abnormal returns after earnings announcements (see, e.g., Bernard and Thomas 1989; Foster et al. 1984; Freeman and Tse 1989; Watts 1978). That is, changes in security prices appear to lag behind changes in earnings expectations as well as changes in earnings.

Section I describes the data and the sample. Section II presents the proxies used to measure the information content of analyst forecast revisions. Section III describes the method of detecting abnormal performance. Section IV reports the main results and section V examines the sensitivity of the main results. Conclusions are in section VI.

I. Databases and Sample Description

Individual analyst forecasts of annual earnings per share (EPS) are obtained from Zacks Investment Research, Inc. for firms with fiscal years ending in 1980–1985. Returns for firms listed on the New York Stock Exchange or American Stock Exchange are provided by the Center for Research in Security Prices (CRSP).

The forecast revisions included in the abnormal returns analysis meet the following criteria: (1) the revision date and the fiscal year-end of the firm are within 1981–1985; (2) stock returns are available on the CRSP daily returns file; (3) the revision date is within 200 trading days of the date of the analyst's prior forecast and within the current fiscal year of the firm; and (4) there are at least two analysts with an outstanding forecast for the firm on the dates of the original forecast and the revision. Of the approximately 3,600 firms on the Zacks database for fiscal year-ends within 1981–1985, 1,465 have revisions that meet the selection criteria. For these firms, 173,620 individual analyst forecast revisions and 66,306 consensus forecast revisions are subjected to empirical tests.

Brokerage houses notify Zacks of analyst revisions via reports that may not have the actual dates of less newsworthy revisions. If a brokerage house does not supply the actual revision date, Zacks uses the date of the brokerage house report. These report dates are always coincidental with or later than the actual revision dates. Zacks does not identify forecasts as having either an actual date or a brokerage house report date.

For the individual analyst revisions, the date on the Zacks database is defined as event day 0. Thus, revisions with brokerage house report dates are actually revised on or prior to day 0. Any immediate abnormal price reaction caused by these revisions will appear on or prior to event day 0. Thus, the use of brokerage house report dates is not an explanation for any abnormal performance after event day 0.

Note that Zacks did not supply the dates that revisions became available to subscribers. Thus, although the revisions are publicly available on or prior to event day 0, any trading strategy must consider the (unknown) cost of gathering revisions.

II. Measures of the Information Content of Forecast Revisions

Individual Analyst Forecast Revisions

A "traditional" measure of information content is:

$$\textit{Traditional Individual } SUF_{i,a,t} = (FRCST_{i,a,t} - FRCST_{i,a,t-v})/FRCST_{i,a,t}, \tag{1}$$

where $SUF_{i,a,t}$ is defined as the scaled unexpected forecast for firm i by analyst a on day t, $FRCST_{i,a,t}$ is analyst a's EPS forecast for company i on day t, a day on which the forecast is revised, and $FRCST_{i,a,t-v}$ is analyst a's prior forecast, which is dated v days prior to day t.[2]

Traditional SUF, which is used by Givoly and Lakonishok (1979, 1980), assumes that $FRCST_{i,a,t-v}$ is the market's expectation of $FRCST_{i,a,t}$ on day $t-1$. However, an efficient market would react to $FRCST_{i,a,t}$ on the basis of expectations on day $t-1$. An "updated" measure that conditions expectations on information available on day $t-1$ is:

$$\textit{Updated Individual } SUF_{i,a,t} = (FRCST_{i,a,t} - E_{t-1}(FRCST_{i,a,t}))/FRCST_{i,a,t}, \tag{2}$$

where $E_{t-1}(FRCST_{i,a,t})$, the expected forecast for day t as of day $t-1$, is estimated as:

$$\begin{aligned} E_{t-1}(FRCST_{i,a,t}) = {} & FRCST_{i,a,t-v} + \hat{\beta}_0 + \hat{\beta}_1(CONSX_{i,t-1} - CONSX_{i,t-v}) \\ & + \hat{\beta}_2(CONSX_{i,t-v} - FRCST_{i,a,t-v}), \end{aligned} \tag{3}$$

and $CONSX_{i,t-1}$ is the consensus forecast, excluding analyst a, for company i on day $t-1$.

Details of this model for estimating $E_{t-1}(FRCST_{i,a,t})$ are in Stickel (1990). Briefly, an individual analyst's next forecast is predicted by updating the analyst's current forecast. The consensus forecast revision between days $t-v$ and $t-1$ proxies for new information released after the date of analyst a's current forecast. The difference between the consensus forecast and analyst a's forecast on day $t-v$ proxies for pressure on analysts to issue forecasts closer to the consensus. Stickel finds a positive relation between each of these two variables and the change in analyst a's forecast between days $t-v$ and t and reports that these two variables explain about 38 percent of forecast revision variability.[3]

---

[2] A divisor of less than \$.20, including any negative number, is arbitrarily set equal to \$.20 to mitigate small denominators. The change in forecast is scaled by the new forecast to have the same denominator as updated SUF, which is defined below. Nevertheless, the results are generally insensitive to using the old forecast, the stock price, or the cross-sectional standard deviation of analyst forecasts at day $t-1$ as the scaling factor.

Parameters are estimated by first segregating revisions into semimonthly subsamples on the basis of day t. Cross-sectional regressions are performed for each subsample, and parameters are averaged across subsamples (see Fama and MacBeth 1973). Parameters are estimated with forecast data from the prior year. Thus, 1980 data is used to estimate parameters for updated 1981 forecasts, and so on.

Consensus Analyst Forecast Revisions

The following measures of the information content of consensus revisions are computed on the fifteenth and last business day of each month (semimonthly).

$$\textit{Traditional Consensus } SUF_{i,t} = (CONS_{i,t} - CONS_{i,t-p})/CONS_{i,t}; \tag{4}$$

$$\textit{Updated Consensus } SUF_{i,t} = (UCONS_{i,t} - UCONS_{i,t-p})/UCONS_{i,t}. \tag{5}$$

$CONS_{i,t}$ is the mean consensus of all outstanding forecasts on day t, with forecasts dated on or prior to day t. This is the same definition of consensus used in prior studies. $UCONS_{i,t}$ is the mean consensus with updated forecasts. Subscript $t-p$ refers to the prior semimonthly period. Traditional $SUF_{i,t}$ is a popular measure of information content (see, e.g., Imhoff and Lobo 1984). Although it measures changing expectations of earnings, it does not use updated forecasts and thus is expected to be a noisier measure.

III. Method of Detecting Abnormal Performance

Measurement of Abnormal Returns

Expected returns are estimated from the market model (see, e.g., Fama 1976). For individual analyst revisions, event day 0 is defined as the Zacks date for a revision. For consensus revisions, day 0 is defined as the end of the semimonthly period over which the change in consensus is measured. Equally weighted CRSP market returns and firm returns from event days +251 to +350 are used to estimate ordinary least squares parameters. A minimum of 30 returns are required for model estimation.[4] The period preceding the event is rejected as the estimation period because analysts are likely to respond to the information in past returns when making a revision (see Copeland and Mayers 1982).

Event Periods

Isolating the impact of forecast revisions is difficult because forecasts are *sequentially* disseminated to market participants by sales personnel in the brokerage house rather than immediately disseminated by major news services, such as the Broad Tape. Sequential dissemination may result in a prolonged price response after event day 0 (see, e.g., Stickel 1985). As reported below, abnormal returns are positively associated with revisions from at least two months prior to revisions until approximately six months after revisions. Statistical significance is reported for the following four cumulation periods:

---

[3] Stickel (1990) also finds a significantly positive relation between the change in the analyst forecast and stock return between days $t-v$ and $t-1$. However, the marginal explanatory power of stock return is negligible.

[4] For the two percent of the sample that did not meet the minimum of 30 returns, market-adjusted returns $(R_{it}-R_{mt})$ are used. The conclusions do not change if these observations are excluded. The results are also robust to estimating market model parameters over days −400 to −301 and estimating abnormal returns as $(R_{it}-R_{mt})-MMAR$, where $MMAR$ is the mean market-adjusted return over days +251 to +350 (see Penman 1987).

Subperiod (−42,0): Day −42 through day 0. Day −42 is about two calendar months before day 0.
Subperiod (+1,+11): Day 11 is, on average, the end of one semimonthly period. Subperiod (+1,+11) is assumed to be the period within which revisions directly affect security prices. Day 0 is excluded because any price movement on day 0 could be caused by other events (such as earnings announcements) that *cause* the analyst forecast revision.
Subperiod (+12,+125): Day 125 is about the end of a six-month holding period.
Subperiod (+126,+250): Day 250 is about the end of a one-year holding period.

Tests of Statistical Significance

The sample is segregated into subsamples on the basis of the semimonthly period in which day t falls, and mean cumulative abnormal returns are calculated by semimonthly subsample. For 1981–1985, there are 120 subsamples (five years × 24 semimonthly periods). This design subsumes cross-sectional temporal dependence *within* subsamples and minimizes dependence *between* subsamples (see Jaffe 1974 and Mandelker 1974).[5] Subsample means are averaged as:

$$MCAR_{(a,b)} = \sum_{s=1}^{P} MCAR_{s,(a,b)}/P, \tag{6}$$

where $MCAR_{s,(a,b)}$ is the mean cumulative abnormal return from event day a to b for semimonthly period s, and P is the number of semimonthly periods. The significance of this overall mean is tested by:

$$t_{(a,b)} = MCAR_{(a,b)}/[\hat{\sigma}_{MCAR_s}/\sqrt{P}], \tag{7}$$

where $\hat{\sigma}_{MCAR_s}$ is the estimated standard deviation of the semimonthly $MCAR$. This test statistic is assumed to be *Student* − t distributed with $P-1$ degrees of freedom.

IV. Empirical Results

Abnormal Returns Surrounding Individual Analyst Revisions

Table 1 reports the mean abnormal returns surrounding individual analyst revisions segregated by magnitude, and separates revisions that are in the top or bottom five percent of the semimonthly distribution of scaled unexpected forecast.[6] This filter

---

[5] As the cumulation period (a,b) lengthens, the calendar overlap across semimonthly periods increases. However, Brown and Warner (1985) find that tests with market model residuals are well-specified in the presence of extreme calendar clustering. Eliminating every other semimonthly period changes no conclusions.

[6] For individual analyst revisions, the magnitude of the revision is compared to the top five percent, bottom five percent, and median cutoffs of the most recently ended semimonthly period. This avoids using distributional information not available at the date of the revision (see Holthausen 1983). Consensus analyst revisions are measured semimonthly, and cutoffs are determined at the end of those semimonthly periods.

Table 1

Abnormal Returns Surrounding Individual Analyst Forecast Revisions with Traditional and Updated Measures of Market Expectations of Forecasts

Individual Analyst Revisions

	Percentile in the Distribution of Traditional and Updated Individual SUF							
	SUF<5th		*5th<SUF<50th*		*50th<SUF<95th*		*95th<SUF*	
	(*Top* 5%)						(*Bottom* 5%)	
	Traditional	*Updated*	*Traditional*	*Updated*	*Traditional*	*Updated*	*Traditional*	*Updated*
$MCAR_{(-42,0)}$	4.60%[a]	0.73%[c]	1.17%[a]	0.88%[b]	−3.43%[a]	−2.69%[a]	−6.33%[a]	−6.10%[a]
t-statistic	(11.88)	(1.77)	(2.86)	(2.32)	(−9.27)	(−6.94)	(−11.39)	(−11.86)
$MCAR_{(+1,+11)}$	0.67%[a]	0.35%[b]	0.19%	0.20%[c]	−0.35%[a]	−0.33%[a]	−0.64%[a]	−0.54%[b]
t-statistic	(4.08)	(2.18)	(1.46)	(1.67)	(−3.03)	(−2.79)	(−2.85)	(−2.55)
$MCAR_{(+12,+125)}$	2.16%[b]	0.56%	1.13%	0.93%	−0.81%	−0.43%	−2.89%[b]	−2.91%[b]
t-statistic	(2.18)	(0.46)	(1.37)	(1.16)	(−0.99)	(−0.55)	(−2.03)	(−2.24)
$MCAR_{(+126,+250)}$	0.13%	−0.92%	0.84%	0.91%	0.31%	0.35%	−1.73%	−1.27%
t-statistic	(0.14)	(−0.85)	(1.08)	(1.20)	(0.40)	(0.45)	(−1.28)	(−1.05)
$MCAR_{(+1,+125)}$	2.83%[a]	0.91%	1.32%	1.13%	−1.16%	−0.76%	−3.53%[b]	−3.45%[b]
t-statistic	(2.66)	(0.71)	(1.48)	(1.30)	(−1.33)	(−0.91)	(−2.33)	(−2.50)
Number of Revisions	9,074	8,978	77,843	77,921	77,540	77,589	9,163	9,132

Total number of individual analyst revisions = 173,620

Revisions are segregated by magnitude and are dated within the current fiscal years of the firms. Fiscal year-ends are within 1981–1985. The *t*-statistics are calculated by using the mean and standard deviation of 120 semimonthly means.

Traditional Individual $SUF_{i,a,t} = (FRCST_{i,a,t} - FRCST_{i,a,t-1})/FRCST_{i,a,t}$;

Updated Individual $SUF_{i,a,t} = (FRCST_{i,a,t} - E_{t-1}(FRCST_{i,a,t}))/FRCST_{i,a,t}$; and

$MCAR_{(a,b)}$ = Mean cumulative abnormal returns from event day *a* to *b*.

[a] Significantly different from zero at less than the 0.01 level.

[b] Significantly different from zero at less than the 0.05 level.

[c] Significantly different from zero at less than the 0.10 level.

is similar to those used by Givoly and Lakonishok (1979, 1980) and Hawkins et al. (1984).

There is evidence that individual analyst revisions, as measured by traditional *SUF*, affect security prices. Table 1 reports that $MCAR_{(+1,+11)}$ is 0.67 percent for the top five percent and −0.64 percent for the bottom five percent. These abnormal returns are significant at less than the 0.01 level.[7] $MCAR_{(+12,+125)}$ suggests that the market slowly assimilates the information in revisions.[8]

Table 1 also reports mean abnormal returns surrounding individual analyst updated *SUF*. Surprisingly, the magnitude of $MCAR_{(+1,+11)}$ is smaller when updated *SUF* is used. With paired comparisons *t*-tests, where mean differences are computed by semimonthly period, the difference between $MCAR_{(+1,+11)}$ for traditional and updated *SUF* is significant at about the 0.01 level for the top five percent (*t*-statistic = 2.56), but insignificant at conventional levels for the bottom five percent (*t*-statistic = −1.15). At least for large upward revisions, investors do not incorporate some publicly available information when reacting to revisions.

Abnormal Returns Surrounding Consensus Analyst Revisions

Table 2 reports results segregated by the magnitude of traditional and updated consensus *SUF*. Again, there is significant abnormal performance *after* revisions.[9] The subsequent abnormal returns are larger with updated consensus *SUF*. With paired comparisons *t*-tests, the difference between $MCAR_{(+1,+11)}$ for updated and traditional *SUF* is significant at less than the 0.01 level for the bottom five percent (*t*-statistic = −2.65), but not significant for the top five percent (*t*-statistic = 0.57). At least for negative revisions, changes in a consensus of updated forecasts are a less noisy predictor of changing prices.

It is important to note the absence of significant abnormal returns over event days +126 to +250. These results suggest that event days +251 to +350 are an unbiased benchmark period and add support for the conclusion that the abnormal returns over event days +1 to +125 are not attributable to a biased estimate of expected return.

Results for an Aggressive Trading Strategy

The individual analyst results are curious because price reaction is more closely associated with the traditional measure of unexpected forecast. A trading strategy based on this finding is to invest in securities with large deviations between the updated and current outstanding forecasts. That is, invest in securities that are predicted to have

---

[7] A large sample size can increase the power of a test to the point where the null is rejected for trivial amounts. The abnormal returns documented in this study are large relative to other capital markets studies and are economically, as well as statistically, significant.

[8] The significant abnormal returns over event days −42 to 0 could be attributable to analysts, responding to the information in price changes, analysts, trading on information prior to issuing the forecast, or, as noted earlier, Zacks, using dates of brokerage house reports.

[9] My traditional *SUF* results help reconcile Givoly and Lakonishok (1979) with Brown et al. (1985). Givoly and Lakonishok (1979) examine the top five percent *SUF*, finding a two-month abnormal return of 2.7 percent. Brown et al. (1985) examine *all positive SUF*, finding an 11-month abnormal return of less than one percent. On table 2, I report a six-month abnormal return of 2.99 percent for the top five percent *SUF* and 1.36 percent when *SUF* is between the fifth and fiftieth percentiles. My results are *inconsistent* with those in Hawkins et al. (1984), which uses the change in the traditional consensus and reports a 12-month abnormal return of 14.2 percent.

Table 2
Abnormal Returns Surrounding Consensus Analyst Forecast Revisions with Traditional and Updated Measures of the Mean Consensus Forecast

Consensus Analyst Revisions

	Percentile in the Distribution of Traditional and Updated Consensus SUF							
	SUF<5th		*5th<SUF<50th*		*50th<SUF<95th*		*95th<SUF*	
	(Top 5%)						*(Bottom 5%)*	
	Traditional	*Updated*	*Traditional*	*Updated*	*Traditional*	*Updated*	*Traditional*	*Updated*
$MCAR_{(-42,0)}$	4.58%^{a}	3.58%^{a}	1.50%^{a}	1.54%^{a}	−2.48%^{a}	−2.38%^{a}	−5.63%^{a}	−6.26%^{a}
t-statistic	(13.17)	(9.98)	(4.10)	(4.48)	(−8.50)	(−7.54)	(−11.07)	(−13.47)
$MCAR_{(+1,+11)}$	0.72%^{a}	0.80%^{a}	0.14%	0.32%^{a}	−0.26%^{b}	−0.40%^{a}	−0.56%^{b}	−0.93%^{a}
t-statistic	(3.94)	(4.80)	(1.12)	(2.64)	(−2.56)	(−3.81)	(−2.38)	(−4.15)
$MCAR_{(+12,+125)}$	2.27%^{a}	3.28%^{a}	1.22%^{c}	1.72%^{b}	−0.69%	−1.21%^{c}	−2.23%	−3.32%^{b}
t-statistic	(2.76)	(3.83)	(1.68)	(2.43)	(−1.08)	(−1.91)	(−1.63)	(−2.54)
$MCAR_{(+126,+250)}$	0.20%	−1.31%	0.68%	0.91%	0.53%	0.38%	−1.07%	−0.66%
t-statistic	(0.26)	(−1.61)	(0.95)	(1.33)	(0.83)	(0.56)	(−0.89)	(−0.54)
$MCAR_{(+1,+125)}$	2.99%^{a}	4.08%^{a}	1.36%^{c}	2.03%^{a}	−0.95%	−1.62%^{b}	−2.80%^{c}	−4.25%^{a}
t-statistic	(3.47)	(4.56)	(1.74)	(2.66)	(−1.38)	(−2.34)	(−1.93)	(−3.07)
Number of Revisions	3,259	3,259	29,897	30,118	29,890	29,665	3,260	3,264

Total number of consensus analyst revisions = 66,306

Revisions are segregated by magnitude and are dated within the current fiscal year of the firm. Fiscal year-ends are within 1981–1985. The *t*-statistics are calculated by using the mean and standard deviation of 120 semimonthly means.

Traditional Consensus $SUF_{i,t}=(CONS_{i,t}-CONS_{i,t-p})/CONS_{i,t}$;

Updated Consensus $SUF_{i,t}=(UCONS_{i,t}-UCONS_{i,t-p})/UCONS_{i,t}$; and

$MCAR_{(a,b)}$ = Mean cumulative abnormal returns from event day *a* to *b*.

[a] Significantly different from zero at less than the 0.01 level.

[b] Significantly different from zero at less than the 0.05 level.

[c] Significantly different from zero at less than the 0.10 level.

large forecast revisions, as measured by traditional individual *SUF*. This strategy is operationalized with consensus forecasts as follows:

$$\textit{Aggressive Consensus } SUF_{i,t} = (UCONS_{i,t} - CONS_{i,t})/UCONS_{i,t}. \qquad (8)$$

Aggressive *SUF* uses updated forecasts to predict the direction of future revisions in current outstanding forecasts. Consensus forecasts are used because the abnormal returns are slightly larger after consensus revisions than after individual analyst revisions.

Table 3 reports abnormal returns after aggressive consensus *SUF*, and figure 1 plots $MCAR_{(+1,+250)}$, restated in terms of approximate semimonthly periods, for securities experiencing a consensus revision in the top and bottom five percent of the semimonthly distributions of traditional, updated, and aggressive consensus *SUF*.[10] Using paired comparisons *t*-tests with mean differences computed semimonthly, I find that $MCAR_{(+1,+125)}$ for the top five percent aggressive *SUF* is significantly greater than that for traditional and for updated consensus *SUF* (*t*-statistics of 5.53 and 4.59, respectively). For the bottom five percent, the difference between aggressive and traditional *SUF* is significant (*t*-statistic $= -3.19$) and the difference between aggressive and updated *SUF* is insignificant (*t*-statistic $= -1.47$).

For comparison purposes, table 3 also reports results when the difference in consensus forecasts is scaled by price per share and the standard deviation of forecasts on day $t-1$. Subsequent abnormal returns are greatest when the difference is scaled by the dispersion in forecasts (see Imhoff and Lobo 1984 for related evidence). For the bottom five percent, there are no significant differences in $MCAR_{(+1,+125)}$ with different divisors. For the top five percent, the scaling factors of forecast and standard deviation of forecasts result in significantly greater $MCAR_{(+1,+125)}$ than scaling by price (*t*-statistics of 2.10 and 2.33, respectively). The difference between using forecast and the standard deviation of forecasts is insignificant at conventional levels.

V. Sensitivity Analyses

Concurrent Earnings, Dividend, and Stock-Split Announcements

Stickel (1989) finds that earnings announcements trigger revisions, which confounds a test of the effect of revisions on prices and further suggests that analysts could be responding to other cues, such as dividend or stock-split announcements, when timing their revisions. From Foster et al. (1984, 588), a large portion (but not all) of the post-earnings-announcement drift in abnormal returns occurs within 11 days after the earnings announcement date. Studies of the effect of dividend and stock-split declarations on stock prices generally find price reaction restricted to event days -1 through $+2$ relative to the declaration. For example, Asquith and Mullins (1983) find reaction on days -1 and 0, and Grinblatt et al. (1984) find reaction on days 0, $+1$, and $+2$.

Table 4 reports $MCAR_{(+1,+11)}$ for samples segregated by whether or not the revision occurs within 11 days of an earnings announcement or within two days of the

---

[10] The abnormal returns for the top five percent aggressive *SUF* exceed those for the bottom five percent in 85 percent of the calendar months within 1981–1985, which indicates that the results are not driven by a few unusual months. $MCAR_{(+1,+125)}$ for securities in the top and bottom five to ten percent of aggressive *SUF* are 5.29 percent (*t*-statistic $= 5.06$) and -0.60 percent (*t*-statistic $= -0.47$), respectively. Thus, the upward revisions are robust to the filter rule, but not the downward revisions.

Table 3

Abnormal Returns for a Trading Strategy that Predicts Changes in Current Outstanding Forecasts

Consensus Analyst Revisions

	Top 5%			*Bottom 5%*		
		Other Scaling Factors			*Other Scaling Factors*	
	Aggressive SUF	*Price*	$\sigma_{forecasts}$	Aggressive SUF	*Price*	$\sigma_{forecasts}$
$MCAR_{(-42,0)}$	5.66%[a]	4.73%[a]	5.82%[a]	−5.56%[a]	−8.04%[a]	−4.17%[a]
t-statistic	(11.74)	(10.52)	(−12.61)	(−10.79)	(−16.45)	(−11.50)
$MCAR_{(+1,+11)}$	1.32%[a]	1.37%[a]	1.49%[a]	−0.95%[a]	−1.00%[a]	−1.34%[a]
t-statistic	(6.95)	(7.25)	(7.80)	(−3.92)	(−3.90)	(−7.99)
$MCAR_{(+12,+125)}$	6.91%[a]	6.20%[a]	7.87%[a]	−4.48%[a]	−5.04%[a]	−5.31%[a]
t-statistic	(7.21)	(6.75)	(9.19)	(−3.06)	(−3.95)	(−7.44)
$MCAR_{(+126,+250)}$	−0.23%	−0.25%	1.11%	−2.04%	−2.15%	0.50%
t-statistic	(−0.32)	(−0.31)	(1.41)	(−1.40)	(−1.69)	(0.74)
$MCAR_{(+1,+125)}$	8.22%[a]	7.57%[a]	9.36%[a]	−5.54[a]	−6.04%[a]	−6.65%[a]
t-statistic	(8.14)	(7.75)	(10.25)	(−3.54)	(−4.44)	(−8.64)
Total Number of Revisions	3,264	3,263	3,252	3,273	3,261	3,276

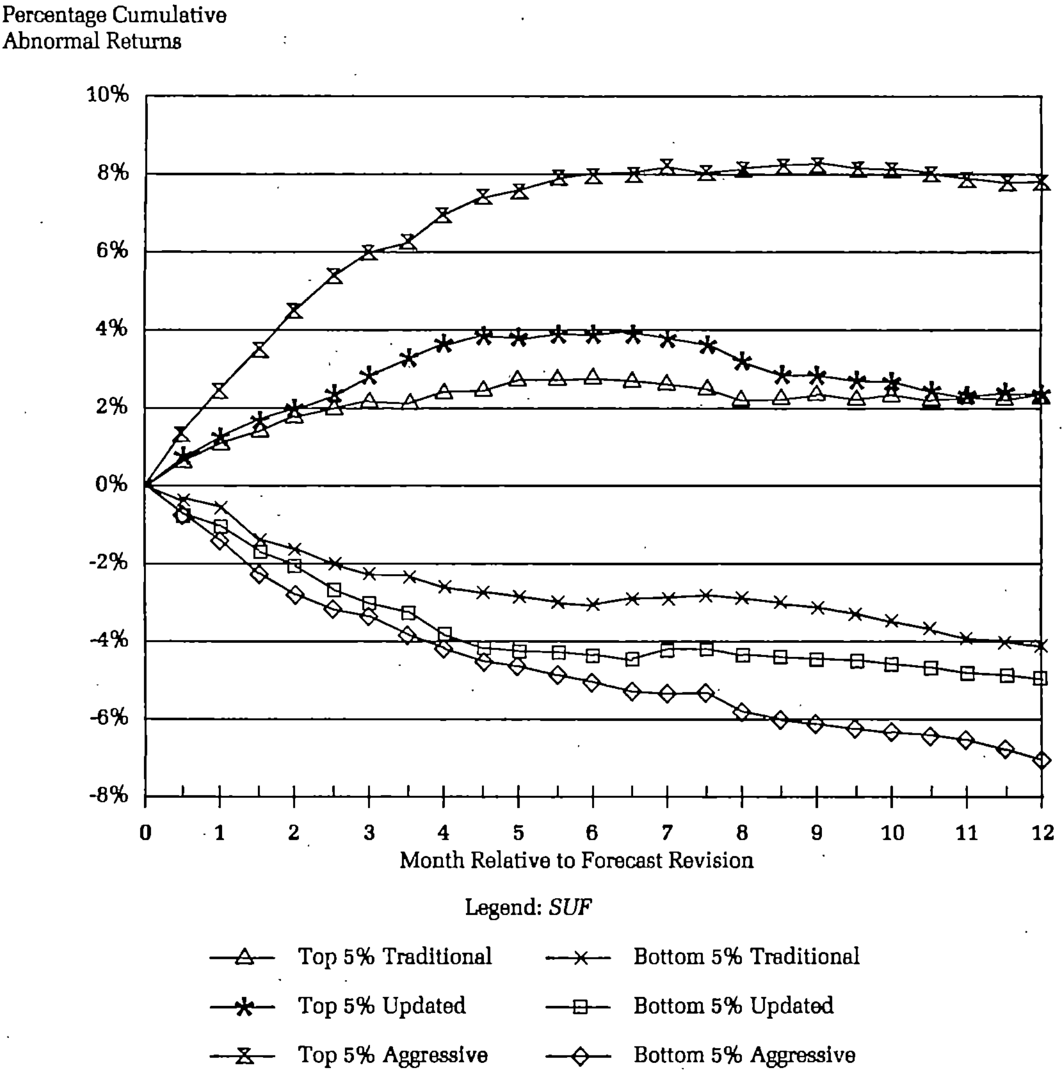
Revisions are dated within 1981–1985 and are in the top or bottom five percent of the semimonthly distributions of scaled unexpected forecast (*SUF*). The *t*-statistics are calculated by using the mean and standard deviation of 120 semimonthly means.

Aggressive Consensus $SUF_{i,t} = (UCONS_{i,t} - CONS_{i,t})/UCONS_{i,t}$; and

$MCAR_{(a,b)}$ = Mean cumulative abnormal returns from event day *a* to *b*.

[a] Significantly different from zero at less than the 0.01 level.

Figure 1
Mean Cumulative Abnormal Returns
After Consensus Forecast Revisions



declaration date of a dividend or stock-split. If revisions affect prices independently of earnings, dividend, and stock-split announcements, the $MCAR_{(+1,+11)}$s should be approximately equal. Again, the analysis is performed for traditional individual *SUF* and updated consensus *SUF*.[11]

---

[11] Dates of earnings announcements are taken from the 1987 COMPUSTAT Quarterly Industrial and Full Coverage Files. Announcement dates for the interim or annual earnings announcement that occurs immediately before and after the forecast revision date are available on COMPUSTAT for approximately 70 percent of the revisions used in the abnormal returns analysis (e.g., for revisions in the top five percent of traditional *SUF*, (3,793+2,794)/9,074=73 percent of the table 1 revisions are included in table 4).

Table 4

Abnormal Returns Following Forecast Revisions with Segregation of Revisions Confounded by Earnings, Dividend, or Stock-Split Announcements

	Traditional Individual SUF			*Updated Consensus SUF*		
	Confounded	*Not Confounded*	*Difference*	*Confounded*	*Not Confounded*	*Difference*
Top 5% SUF:						
$MCAR_{(+1,+11)}$	1.05%[a]	0.85%[a]	0.20%	1.05%[a]	0.84%[a]	0.21%
t-statistic	(4.06)	(3.02)	(0.51)	(2.68)	(2.74)	(0.43)
Number of Revisions	3,793	2,794		974	1,086	
Bottom 5% SUF:						
$MCAR_{(+1,+11)}$	−0.55%	−0.67%[b]	0.12%	−1.32%[a]	−0.81%[b]	−0.51%
t-statistic	(−1.54)	(−1.93)	(0.23)	(−3.24)	(−1.74)	(−0.82)
Number of Revisions	3,503	3,088		1,004	1,048	

Confounded revisions occur within 11 trading days of an earnings announcement or within two days of a dividend or stock-split declaration. Revisions are in the top or bottom five percent of the semimonthly distributions of *SUF*. The *t*-statistics for *MCAR* following confounded and not confounded revisions are calculated by using the mean and standard deviation of 120 semimonthly means. The *t*-statistics for the difference in *MCAR* are for a test of difference in mean that uses semimonthly means and assumes unequal variances.

$MCAR_{(+1,+11)}$ = Mean cumulative abnormal returns from event day +1 to +11.

[a] Significantly different from zero at less than the 0.01 level.

[b] Significantly different from zero at less than the 0.10 level.

There is no significant difference between the abnormal returns following revisions that are confounded and those not confounded. This suggests that there is information in revisions beyond that in earnings, dividend, and stock-split announcements.

Risk Changes

An implication of Ball et al. (1988) is that beta shifts could explain the abnormal returns subsequent to forecast revisions. Estimating betas for a 60-day period on each side of event day 0 for firms in the top and bottom five percent of the semimonthly distributions of traditional individual *SUF* yields a slight increase in mean beta, from 1.37 to 1.39, after positive revisions. After negative revisions, the mean beta decreases from 1.38 to 1.37.[12] As in Bernard and Thomas (1989), these beta shifts are too small to explain subsequent abnormal performance.

VI. Conclusions

Earnings forecast revisions affect stock prices, and this price effect is a function of the magnitude of the revision. Nevertheless, prices are slow to assimilate the information. These results cannot be explained by changes in beta risk or by an association between forecast revisions and the post-earnings-announcement drift.[13] Surprisingly, abnormal returns following revisions continue to drift in the direction of the revision for about six months.

Stickel (1990) finds that updated forecasts are less biased and more accurate predictors of future forecasts than current outstanding forecasts. Surprisingly, stock returns immediately after individual-analyst revisions suggest that the current outstanding forecast is a better measure of market expectations. This finding suggests an aggressive trading strategy that uses updated forecasts to predict large revisions of current outstanding forecasts. The abnormal returns for securities predicted to perform best and worst average 8.22 and −5.44 percent, respectively, every *six months.*

The process by which changing expectations of earnings are assimilated into prices is odd. The inconsistency of these results with an efficient capital market may be attributable to trading costs and the costs of gathering and processing analyst forecasts. Thus, although prices appear inefficient because they do not immediately assimilate the information in forecasts, no claim is made that an investor could earn positive abnormal returns after these costs. Although reasonable estimates of transaction costs are available, the costs of gathering and processing forecasts are much more difficult to approximate.

---

[12] Finding a mean beta greater than 1.0 is expected because firms in the top and bottom five percent of *SUF* are, on average, smaller firms. Betas are also calculated by cumulating individual security returns and cumulating market returns for each 60-day period, which produces a single observation for a cross-sectional regression analogous to Ibbotson (1975). These beta estimates are closer to 1.0, which is unexpected and unexplained. Nevertheless, the changes in these betas are again incapable of explaining the abnormal returns after forecast revisions.

[13] Atiase (1985) finds that smaller firms have a larger price reaction to earnings announcements. An interesting question for future research is whether price reaction to forecast revisions is a function of firm size.

References

Abdel-khalik, A. R., and B. Ajinkya. 1982. Returns to informational advantages: The case of analysts' forecast revisions. *The Accounting Review* 57 (October): 661–80.

Asquith, P., and D. W. Mullins. 1983. The impact of initiating dividend payments on shareholders' wealth. *Journal of Business* 56 (January): 77–96.

Atiase, R. K. 1985. Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research* 23 (Spring): 21–36.

Ball, R., S. P. Kothari, and R. L. Watts. 1988. The economics of the relation between earnings changes and stock returns. Working paper, University of Rochester.

Bernard, V. L., and J. K. Thomas. 1989. Post-earnings-announcement drift: Delayed price response, or risk premium? *Journal of Accounting Research* 27 (Supplement): 1–36.

Brown, P., G. Foster, and E. Noreen. 1985. *Security Analyst Multi-Year Earnings Forecasts and the Capital Market.* Sarasota, FL: American Accounting Association.

Brown, S. J., and J. B. Warner. 1985. Using daily stock returns: The case of event studies. *Journal of Financial Economics* 14 (March): 3–31.

Copeland, T. E., and D. Mayers. 1982. The value line enigma (1965–1978): A case study of performance evaluation issues. *Journal of Financial Economics* 10 (November): 289–321.

Fama, E. F. 1976. *Foundations of Finance.* New York: Basic Books.

——, and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81 (May/June): 607–36.

Foster, G., C. Olsen, and T. Shelvin. 1984. Earnings releases, anomalies, and the behavior of security returns. *The Accounting Review* 59 (October): 574–603.

Freeman, R. N., and S. Tse. 1989. The multiperiod information content of earnings announcements: Rational delayed reactions to earnings news. *Journal of Accounting Research* 27 (Supplement): 49–79.

Givoly, D., and J. Lakonishok. 1979. The information content of financial analysts' forecasts of earnings: Some evidence on semi-strong inefficiency. *Journal of Accounting & Economics* 1 (March): 165–85.

——, and ——. 1980. Financial analysts' forecasts of earnings: Their value to investors. *Journal of Banking and Finance* 4 (September): 221–34.

Grinblatt, M. S., R. W. Masulis, and S. Titman. 1984. The valuation effects of stock splits and stock dividends. *Journal of Financial Economics* 13 (December): 461–90.

Hawkins, E. H., S. C. Chamberlin, and W. E. Daniel. 1984. Earnings expectations and security prices. *Financial Analysts Journal* 40 (September/October): 24–27; 30–38; 74.

Holthausen, R. 1983. Abnormal returns following quarterly earnings announcements. *Proceedings of the Seminar on the Analysis of Security Prices*, University of Chicago: 37–59.

Ibbotson, R. G. 1975. Price performance of common stock new issues. *Journal of Financial Economics* 2 (September): 235–72.

Imhoff, E. A., and G. J. Lobo. 1984. Information content of analysts' composite forecast revisions. *Journal of Accounting Research* 22 (Autumn): 541–54.

Jaffe, J. 1974. The effect of regulation changes on insider trading. *Bell Journal of Economics and Management Science* 5 (Spring): 92–121.

Lys, T., and S. Sohn. 1989. The association between revisions of financial analysts' earnings forecasts and security prices. Working paper, Northwestern University.

Mandelker, G. 1974. Risk and return: The case of merging firms. *Journal of Financial Economics* 1 (December): 303–35.

Penman, S. H. 1987. The distribution of earnings news over time and seasonalities in aggregate stock returns. *Journal of Financial Economics* 18 (June): 199–228.

Stickel, S. E. 1985. The effect of value line investment survey rank changes on common stock prices. *Journal of Financial Economics* 14 (March): 121–43.

——. 1989. The timing of and incentives for annual earnings forecasts near interim earnings announcements. *Journal of Accounting & Economics* 11 (July): 275–92.

——. 1990. Predicting individual analyst earnings forecasts. *Journal of Accounting Research* 28 (Autumn): 409–17.

Watts, R. L. 1978. Systematic abnormal returns after quarterly earnings announcements. *Journal of Financial Economics* 6 (June/September): 127–50.

THE ACCOUNTING REVIEW
Vol. 66, No. 2
April 1991
pp. 417–428

BOOK REVIEWS

Shane R. Moriarity, Editor

Editor's Note: Two copies of books for review should be sent to Professor Shane Moriarity, School of Accounting, University of Oklahoma, Norman, OK 73019. The policy of the *Review* is to publish only those reviews solicited by the Book Review Editor. Unsolicited reviews will not be accepted.

SIMON ARCHER, STUART McLEAY, and JEAN-BRUNO DuFOUR, *Audit Reports on the Financial Statements of European Multinational Companies: A Comparative Study* (London: Institute of Chartered Accountants in England and Wales, 1989, pp. 99, £10.00, paper).

This is a report of a study commissioned by the Research Board of the Institute of Chartered Accountants in England and Wales. It is one part of a more extensive research program on financial reporting by European companies that is organized by the first two authors, both at the University of Wales in Bangor. It is motivated by the concern that national differences in financial reporting practices may give rise to a type of information asymmetry which is likely to put nonnational users at a disadvantage compared to the residents of the country in which the financial report originates. Furthermore, this condition of information asymmetry may be aggravated because nonnational users may have less understanding than nationals of the degree and nature of the audit assurance given in the annual report. Others share this concern as two international bodies have been active in efforts to harmonize audit practices of different nations. The International Auditing Practice Committee (IAPC) of the International Federation of Accountants (IFAC) has been setting international auditing guidelines (IAG) for worldwide application, and the Auditing Statement Board (ASB) of the Union Europeene des Experts Comptables Economiques et Financiers (UEC) has been doing the same in a European context. Both the IAPC and the ASB issued guidelines on the auditor's report; IAG No. 13 *The Auditor's Report on Financial Statements* (1983) and ASB No. 8 *The Audit Report* (May 1981).

The current study compares the audit reports of European companies with multiple listing on European exchanges using criteria indicated in IAG 13 and ASB 8, giving particular emphasis to criteria considered helpful to international readers. The sample is made of 206 companies out of a total of 245 companies that have multiple listings from 16 European countries (ten of which are members of the European Community). [The sixteenth country, Liechtenstein, with a sample of one report was omitted from the analysis.] Twenty-one characteristics were used to compare the auditors' reports in the annual reports of these companies to IAG 13 and to a specimen report prepared by each country for ASB. For reports issued in more than one language, all language versions were analyzed. The study also included a brief description of institutional context in each country based on Price Waterhouse publications *Doing Business.*

An index to measure variability in audit reports within countries was constructed. It indicated a range of variability from none in the U.K. to high variability in Luxembourg. The degree of compliance with IAG 13 was measured in ten items showing a high degree of compliance by Ireland and the U.K. and low compliance by Germany, Sweden, and Austria. The EC members in general have a higher degree of compliance; the authors attribute this to the influence of the Fourth Directive. The authors also attributed differences between countries in compliance with IAG 13 to cultural differences. However, no empirical evidence of the role of cultural differences was presented. While one of the motivating factors for the study is information asymmetry between nationals and nonnationals, no attempt was made to link the findings to this theoretical underpinning. One hopes that the rest of the research program includes an analysis of nondomestic markets' reaction to differences in compliance with IAG 13, the source of information assymetry.

The study includes a wealth of data and it is one of the few studies in international auditing by English speaking researchers that analyzes a significant number of reports in other languages. It would have been helpful if a list of the companies in the sample was included. The study may be useful to nonacademics who are interested in the quality of audit reports of the countries in the samples. For academics, the study can be a source of data to test hypotheses and theories on international auditing.

MOHAMED E. A. HUSSEIN
Associate Professor of Accounting

The University of Connecticut

GRAHAM BANNOCK, *Taxation in the European Community: The Small Business Perspective* (London: Paul Chapman, Ltd., 1990, pp. ix, 130, $39.95, distributed in the U.S. and Canada by Marcus Wiener Publishing, New York).

As 1992 approaches interest in taxation within the European Community (EC) has accelerated. While the tax systems of the individual countries in the EC will remain largely intact, significant changes have been proposed as a result of the imminent removal of fiscal frontiers between member nations. This timely and well written book both examines and critiques the taxation, and certain proposed changes in the taxation of small and medium sized enterprises (SMEs) within the EC.

The book is divided into two distinct sections. The first section provides an overview of the tax systems within the EC relevant to business, and an assessment of whether such discriminate against SMEs. The second section examines the compliance burden imposed upon European businesses by value added taxes (VATs) and the mechanics and implications of proposed changes regarding VATs and excise taxes that would be effective after 1992.

In keeping with its stated objectives the book focuses upon the big picture, rather than providing a detailed description of the tax systems in the EC affecting SMEs. While the reader will gain some appreciation of the basic tax systems affecting business in the EC, the book's primary strength lies in the author's assessment of existing laws and proposals, and recommendations for reform. For example, in a particularly thought provoking part of the book the author argues that the tax and economic systems within the EC unfairly discriminate against SMEs. Factors cited in support of this argument include:

1. Favoritism in the tax law shown certain industries which are dominated by larger firms;
2. The increased burden placed on SMEs by taxes that are flat in amount;
3. The denial of tax deductions, allowances, and exemptions to many owners of SMEs due to their large percentage of ownership in the business; and, most importantly
4. The relatively heavier burden compliance with the tax law places on SMEs.

Several possible reforms are suggested for reducing this alleged discrimination including the abolition of a corporate tax and establishment of a Uniform Expenditure Tax.

Although well reasoned and instructive, the utility of the first section is somewhat hampered by reliance on data often running only through 1984, and derived from a 1986 study initiative of the ESME Contact Group.

The second section provides a thorough description of proposed changes within the EC for 1992 concerning indirect taxes (VATs and excise duties). This section is based upon a 1988 study initiative of the ESME Contact Group. Major proposals discussed include:

1. The elimination of a zero tax rate on exports;
2. The establishment of but two VAT rates;

3. The establishment of a supranational organization to oversee the VATs on imports and exports between member nations, and assure that no member nation runs a net surplus or deficit on VATs from imports and exports; and
4. The harmonization of excise taxes.

The author makes a sound analysis of the problems that the adoption of the proposals will cause from a compliance, auditing, and enforcement perspective.

Despite certain limitations, this is a valuable book. Not only does the book provide the reader with an introduction to taxation within the EC, and insight into significant proposed changes, but perhaps more importantly makes the reader reflect on many aspects and issues of tax policy all too often ignored.

MARK A. SEGAL
Assistant Professor of Accounting
University of South Alabama

JOYCE BERG, LANE DALEY, FRANK GIGLER, and CHANDRA KANODIA, *The Value of Communication in Agency Contracts: Theory and Experimental Evidence* (Vancouver: The Canadian Certified General Accountants' Research Foundation, 1990, pp. ix, 116, CAD $12.00, paper).

This book presents the results of laboratory-market experiments designed to test agency theory predictions regarding communication of private, post-decision information. The project focuses strictly on one possible value of communication in agencies—improved risk sharing. For tractability purposes the study ignores other sources of communication value. (See Baiman (1989) for a discussion of other sources of value of communication.) The economic model consists of a principal who owns a stochastic production technology and an agent who contracts with the principal to provide an unobservable, costly input. After the agent selects and provides a level of input, he privately observes a signal that is imperfectly correlated with the future output and then reports on the signal to the principal. This report is subject to strategic manipulation; it need not be truthful. The principal's problem is to design a contract that will induce the agent to select the most efficient level of input. Agency theory predicts that, ignoring communication costs, the principal will choose a contract involving the communication of private information because this contract will result in welfare improvement (see Dye 1985). The experiments reported here investigate the predictive power of this theory.

The actual tasks performed by the subjects were quite different in appearance than the contractual language might suggest. Each subjects' payoff was determined by spinning a prize wheel which theoretically controlled the subjects' risk preferences. The probability of winning a monetary prize depended both on the subject's action choice and a random outcome. The principal's task was to select one of two prize wheels (representing contracts) for the agent to use. The agent then selected one of two different die to roll (representing his action choice) and whether or not to truthfully report the die roll results. The win probabilities for both players depend on the agent's choice of die (such that higher action is more costly to the agent), a random outcome and, for some wheels, on the agent's report.

The authors conduct two experiments with different reward levels for the principal. Specifically, the difference between the principal's expected payoffs under the optimal contract versus the non-optimal contract is much higher in the second experiment relative to the first. Both experiments indicate that agents respond to different contracts largely as predicted by the theory. That is, agents report (un)truthfully when contracts are designed to induce (non)truthful reporting. Furthermore, agents generally provide the level of input that is optimal given the contract offered them. Principals, on the other hand, do not necessarily offer optimal contracts. In the first experiment the principal often offered the non-optimal contract. In the second experiment, on the other hand, principals chose the optimal contract much more often.

The results of the second experiment seem to support the theory; it is not clear, however, how much we can really learn from this experiment. Is it surprising that when the cost of making a non-optimal choice is sufficiently high (say, being shot in the head) choices will be optimal? We find the less intuitive results of the first experiment more interesting, and will focus our discussion there.

While we might interpret the results of the first experiment as suggesting that the theory has limited predictive power, we can also think of an alternative interpretation for the principals' non-optimal performance. A critical feature affecting a principal's contract choice is his or her belief about the agent's action and reporting choices, and the authors do a good job of revealing these choices by having the principals act as agents in a preexperiment training game. Unfortunately, the authors may have lost control of this belief by fixing the principal-agent pairings for the actual game. As the game proceeds the principal can update his beliefs about the agent's actions, and the agent, knowing that he can affect the principal's beliefs, may attempt to exploit the multi-period nature of their relationship. (The authors appeal to the theoretical concept of backward induction to rule out such behavior, but the predictive power of this concept is an open empirical question.) The study could have avoided these multi-period problems by having each principal play each agent only once.

The principals' failure to behave optimally in the first experiment is particularly surprising given that the agent behavior is very close to the theoretical prediction. A strong follow-up experiment would have been to automate the agents and study the principal's individual decision-making problem. If the principal cannot make the optimal choice in an environment stripped of strategic interactions, then the discrepancy between the theory and the results of the study's first experiment could be pinned to a very specific cause. Contrasting the first experiment with an individual decision-making experiment would enable us to assess whether the seemingly suboptimal behavior on the part of the principals is due to strategic interaction with agents or due to suboptimal decision-making on the part of principals—possibly due to the complexity of task.

The small number of contracts available to the principals might seem like a drastic departure from the theoretical model studied. There is, however, one important advantage in this construction not articulated by the authors. Recall that the Revelation Principle does *not* predict truthful reporting; it only implies that the performance attained by any optimal contract can be replicated via a truth-inducing contract. The experimental design takes care of this problem by only allowing a single optimal communication-based contract. There is, however, a remaining conceptual concern. Agency theory is silent about the equilibrium contract that is likely to emerge. In fact, we have no conclusive experimental evidence supporting the conjecture that truth-telling constitutes a focal point. Therefore, what the authors are actually capturing in the first experiment may be a consequence of the fact that the principal is not allowed to choose the nonincentive-compatible optimal contract. In fact, casual observations suggest that contracts often lead to untruthful communication of information.

Despite the above concerns, the experiments are very well done. The study focuses on a single issue and treats it carefully. The link between the theory and the experimental environment is strong; more so than in many other similar experiments. Many noneconomic factors that could influence the results were controlled and the presentation of both the theory and the results is clear. This work deserves the attention of any researcher interested in the value of communication from either a theoretical or an experimental angle.

RUSSELL LUNDHOLM
Assistant Professor of Accounting
NAHUM MELUMAD
Associate Professor of Accounting

Stanford University

CANADIAN INSTITUTE OF CHARTERED ACCOUNTANTS, *Accounting and Reporting for Physical Assets by Governments* (Toronto: CICA, 1989, pp. xiii, 114, CAD $25.00).

The debate continues on how to account for and report physical assets by governments. Since governments do not need to measure operating income or compute taxable income, the traditional approaches used in the private sector to account and report on physical assets have not been applied to governmental funds. Yet, pressure continues to build for governmental funds to adopt full accrual accounting methods and apply depreciation accounting. Time will tell as to whether these applications will extend to physical assets owned by governments.

The Government Accounting Standards Board (GASB) is presently addressing this same subject in response to their *Discussion Memorandum: An Analysis of Issues Related to Accounting and Financial Reporting for Capital Assets of Governmental Entities* (Stamford, CT: GASB, August 1987). Some pertinent thoughts relative to that discussion memorandum on the issue of excluding depreciation are discussed by Attmore, Miller, and Fountain ("Governmental Capital Assets: The Challenge to Report Decision-Useful Information," *Government Finance Review*, Volume 5, Number 4, August 1989, pp. 13–17). An exposure draft on the proposed GASB standard is expected early in 1991.

This research study by a Canadian Study Group looks at the international dimension of the problems and makes pertinent recommendations for the federal, provincial, and territorial governments in Canada. Recommendations that affect two areas debated extensively in the literature are as follows: (1) the full accrual basis of accounting should be used to account for and report physical assets (fixed assets and inventories), and (2) depreciable assets should be amortized over their useful life and deferred maintenance should be disclosed in a note to the financial statements. To implement the recommendations, some sample financial statements were presented that were very informative.

The Study Group makes a persuasive argument that the full accrual basis of accounting should be used to provide users of financial statements with the information that they need to hold a governing body accountable for its *performance*. They hold that this basis is necessary to provide the cost information critical to decisions relating to the stockage, acquisition, and consumption of physical assets. Further, full accrual will provide the information necessary to assess whether current taxpayers have paid an equitable share of the cost of the benefits provided to them by the government.

A major concern is that existing government accounting practices encourage governments to defer essential maintenance and replacement of physical assets. The Study Group acknowledges the difficulty in identifying the extent of deferred maintenance and recommends the development of innovative approaches to resolve this problem. One solution suggested is "renewals accounting" which was used in the U.S. by the railroad industry. This solution is also suggested by the New Zealand Society of Accountants. This approach assumes that larger infrastructures mature to a steady state where the annual cost of maintenance closely parallels the wearing away of the asset value. Consequently, the actual expenditure for renewal is charged as an expense during the period rather than depreciation.

A major flaw in the recommendations of the Study Group relative to depreciation is the lack of a persuasive argument on the increased value of information that would be provided to decision makers. The suggested use of "renewals accounting" with an allowance for renewals may be a compromise solution that could move us forward in our attempt to better account for infrastructures by governments. However, the "renewals accounting" method (also known as retirement-replacement-betterment) was discontinued by U.S. railroads in 1983 and replaced with depreciation accounting at the direction of the Interstate Commerce Commission since it could permit smoothing of income.

In spite of these shortcomings, the recommendations of the Study Group are a valuable contribution to the continuing debate. Whether one agrees or not with their recommendations, this study highlights the international dimension of the problem. Even if the nations debating the issues reach different solutions, experimentation with different methods may permit the accounting profession to assess the value of additional information for decision making by governments relative to infrastructures.

JESSE W. HUGHES
Associate Professor of Accounting

Old Dominion University

PETER D. CHANT, *The Financial Statement Presentation of Corporate Financing Activities* (Toronto: The Canadian Institute of Chartered Accountants, 1990, pp. xix, 228, CAD $32.50).

Debt or equity? Recognize or disclose? These questions recur regularly in conjunction with the current proliferation of innovative financial instruments. These instruments include: warrants and options, interest rate swaps, participating mortgages, commodity-indexed notes, convertible securities, and so on. In recent years, the creation of such instruments has clearly outpaced the accounting standard setting process. Currently, many instruments are not addressed by accounting standards. Instead, accounting is by analogy or in reference to very dated guidance.

This research monograph, commissioned by the Canadian Institute of Chartered Accountants, is "intended to stimulate thought, discussion, and debate" (p. iv) on the reporting of corporate financing activities. The study begins from the premise that the traditional accounting definitions of debt and equity are no longer adequate and have led to accounting standards that are unclear, conflicting, and lacking in conceptual support. The study then purports to establish a theoretical framework for the reporting of corporate financing activities (Chapters 4 and 5) and to apply this framework with a series of recommendations (Chapters 6–11) for the recognition, classification, and disclosure of financial instruments.

The proposed theoretical framework relies on what the author calls the "corporate governance literature." This is, in fact, a combination of key writings from the agency theory, positive accounting theory, and option pricing research. The author uses this literature to emphasize the conflicting interests of creditors, shareholders, and management. He then concludes that the primary objective of financial reporting should be "to provide information to investors and creditors of an enterprise for purposes of monitoring and motivating the behaviour of management..." (p. 63). Thus, the author has taken the rather bold step of proposing to move the objective of financial reporting away from the prediction of future cash flows (FASB Concepts Statement No. 1) toward an accountability/motivational model.

The proposed framework then serves as a basis for a series of reporting recommendations for nonfinancial entities. (Financial institutions are discussed briefly in a separate chapter.) The recognition criteria take a conservative approach to the identified objective and recommend the reporting of "those actions and events that measurably increase the uncertainty associated with creditors or adversely affect equity interests" (p. 80). This implies the recognition of all adverse consequences (rather than expected outcomes) if measurable and more than remotely possible. The reporting of these consequences would not, however, emphasize discounted present value or fair value concepts. Instead, historical proceeds and contractually-defined amounts would be used to signal possible outcomes. In general, many of the specific reporting rules proposed by the author are consistent with current practice (e.g., purchase commitments are recognized only in the event of "adverse consequences"). Some of the primary departures include: capitalization of all leases that involve repayment of capital, broader recognition of the obligations associated with employee stock options, and stricter guidance for the offsetting of assets and liabilities.

This monograph clearly provides considerable "food for thought." It has accomplished its purpose of generating many ideas for discussion and debate. It has, however, also attempted an immense task. This task includes reviewing the current financing and reporting environments, redefining the objective of financial reporting, and proposing a number of recognition, classification, and disclosure principles. The sheer magnitude of the undertaking implies gaps in the analysis and a clear need for a more precise bridging of the general principles to the reporting of specific financial instruments. The appendices are a good start in this direction. They contain useful descriptions of existing financial instruments, current professional guidance, and applications of the author's proposals to selected financing activities. Clearly, however, further discussion is needed. As such, this monograph is recommended for anyone with an interest in liability/equity recognition and disclosure. It would also be an appropriate text (in conjunction with the FASB's discussion documents on financial instruments) for a graduate-level financial theory course.

ROSANNE M. MOHR
Associate Professor of Accounting

Portland State University

FRANK A. COWELL, *Cheating the Government: The Economics of Tax Evasion* (Cambridge, MA: MIT Press, 1990, pp. xii, 267, $19.95).

Tax compliance has received increased attention from accounting researchers who have sought to understand and explain the determinants of noncompliant behavior. This book provides a significant contribution to this field by establishing a comprehensive theoretical framework for the economics of tax evasion. Cowell develops this theory using microeconomic analysis based on the standard assumptions about rational decision making in the face of risk.

Initially focusing on a taxpayer with exogenous income who is prepared to evade if it is perceived to be financially worthwhile, Cowell demonstrates that people with higher risk aversion tend to evade less, and individuals with higher income tend to evade more. Both of these conclusions are consistent with empirical research. Cowell notes that another conclusion is contrary to prior findings. He analytically demonstrates that increases in tax rates reduce tax evasion. However, when income is considered to be endogenous and determined by the amount of labor a person supplies, tax evasion increases with increases in tax rates. This is because at higher tax rates the proportion of time spent in the illegal (underground) sector will increase. Subsequent analysis incorporates other factors, such as perceived inequity, communal behavior, and the impact of moonlighters and ghosts (people for whom there is no record).

The final part of the book is devoted to policy design. I was very impressed that Cowell did not create these policy recommendations in a vacuum but instead considered the impact of several legal, political, and informational constraints. An interesting part of this section is an analysis of how evasion policy would be affected by the legal alternative to tax evasion, tax avoidance. He demonstrates that tax collection may not increase if tax authorities tighten up on the enforcement of tax law. Some taxpayers will switch from evasion to avoidance rather than pay more taxes.

I found the book to be very well written. Cowell intersperses his formal analysis with anecdotes to illustrate various points. My favorite, in the chapter on morality and community, was Adlai Stevenson's statement, "A lie is an abomination unto the Lord, and a very present help in trouble" (p. 101).

In addition to developing a theoretical framework, Cowell briefly reviews empirical studies in tax compliance. This survey is supplemented with an extensive bibliography. Almost 400 studies are referenced in the bibliography.

In summary, I found this book to be an excellent analysis of the problem of tax evasion. It provides empirical researchers with many testable hypotheses that are based on economic theory. However, not only academics should read the book. Anyone interested in tax evasion and equipped with basic microeconomic skills should find the book interesting.

ROBERT J. ROLFE
Associate Professor of Accounting
University of South Carolina

ALEX EDWARDS and N. A. D. CONNELL, *Expert Systems in Accounting* (Hertfordshire, UK: Prentice-Hall International, 1989, pp. xiv, 254, $54.00).

This is an easy-to-read book which is intended to accomplish the following two objectives:

1. Provide an introduction to decision-support systems and expert systems, and
2. To provide the reader with a sense of the process of expert-system development by presenting a case study of prototype development.

The book contains two sections, one devoted to each of the above objectives. The entire text is written in tutorial style and is well-suited to the reader who is newly interested in expert systems, as well as those who have only a passing interest, but would like to know what all the fuss is about.

The authors also include as original material the results of a mail survey on the extent of involvement of professional accountants in the development of expert systems. This is incorporated with summary information on expert systems under development in the accounting arena.

Part I includes commentary on the definition and role of expert systems—with a particular focus on accounting. The authors are even brave enough to try to clarify the distinction between expert systems and decision-support systems. Since ambiguous terminology seems to run amok in new technologies, this is a worthy effort. The fundamentals of expert systems are presented in a nonthreatening manner, followed by a discussion of expert system R&D in accounting. Part I concludes with an assessment of future directions which may be of interest to both researchers and practitioners.

Part II is devoted to illustrating the process of developing an expert system prototype. This material includes methodologies, pitfalls to be aware of, the characteristics of expert systems prototypes, and tools which can be of use. For the uninitiated, Part II brings to life a number of the concepts discussed in Part I.

In summary, this is a very readable introduction to expert systems and their applications to accounting problems. It won't make the reader an expert on expert systems, but it does provide a clear account of some fundamental concepts.

JAMES V. HANSEN
Professor of Accountancy

Brigham Young University

BELVERD E. NEEDLES, JR. and VERNON K. ZIMMERMAN, Editors, *Comparative International Accounting Educational Standards* (Champaign, IL: University of Illinois Center for International Education and Research in Accounting, 1990, pp. iv, 243, $10.00).

This monograph provides an abundance of information about international accounting education. Individuals wishing to discuss and consider changes in accounting education will find many useful ideas and references. These ideas arise from the different practice requirements, varied educational prerequisites, and the diverse admissions criteria shown to exist internationally in accounting. Papers included address philosophical education matters and practitioner perceptions of accounting education. It is possible to contrast current states of accounting education in developed versus underdeveloped countries. The authors in this monograph also present educational distinctions within varied social systems and across geographic regions. A useful framework for reading the monograph contents exists in the opening paper. However, the usefulness of this opening paper as a guide for readers is not specified or immediately apparent.

The opening paper develops a background and framework for understanding various models of accounting education and entry into the profession. Educational models include the integrated, linear, and parallel forms. The integrated form uses a concurrent exposure to education and experience while the parallel model shows the use of either education or experience leading to professional status. On the other hand, the linear model requires education before experience. Different, and widely varied, types of examinations and testing levels are used within each form. These models show the differences that later become apparent in other monograph contributions. This opening paper also presents the International Accounting Education Guidelines. Although not mentioned, guidelines presented in this first paper help the reader with comparisons in following works.

Several approaches are possible for reading and using the monograph material. These approaches take a general matrix form of developed versus underdeveloped countries, philosophical outlook versus what exists, and centralized versus decentralized social systems. Another further comparison is one geographic region versus another. Many other matrices are readily apparent and papers in the monograph provide sufficient information for study. Using these matrices as a reading guide led to one feeling that centralized educational control is a failure. Another difference is the extensive use of a formal thesis in educational settings and an oral defense in professional examinations.

The major *caveat* which relates to the information is the age of the data. However, most authors did not foresee changes happening quickly in their respective domains. If one wants merely a textbook on international accounting education, this monograph would most likely not be of interest. Several papers are hard to follow due to unfamiliar terms and acronyms. Such difficulty is overcome in most cases by the content of the material and use of the education models. The monograph does not have a set pattern and one should read the papers from a personal or particular view. Sorting papers by desired matrix area simplifies the reading process. Should a reader be interested in a diverse sampling of international accounting education, this work will be of interest. Using the guide provided in the opening paper is important.

THOMAS B. CLEVENGER
Assistant Professor of Accounting
Washburn University of Topeka

CHRISTOPHER NOBES, *Interpreting European Financial Statements: Towards 1992* (London: Butterworths & Co. Ltd., 1989, pp. x, 182, £50.00).

At midnight on December 31, 1992, the European Community hopes to be able to shed most of the barriers to trade and become a truly unified common market. A part of that process involves the free flow of capital, which implies that financial statement users must be able to understand the financial statements of the 12 members of the EC in order to compare financial information and thus make wise investment decisions. The purpose of *Towards 1992* is to help readers, both within and without the EC, understand the similarities and differences in the accounting environments of the member nations of the EC.

The book is written primarily for business practitioners, such as financial managers and auditors of multinational enterprises, as well as financial analysts, investors, and creditors. The book does not contain footnotes, and end references are kept at a minimum.

There are two ways to look at accounting in different countries. The more conventional way is to look at accounting as it fits in the educational, cultural, legal-political, and economic environments of a country. Nobes and Parker did that in *Comparative International Accounting* (1985). In some cases, the descriptive approach becomes comparative with the use of a particular country as a benchmark. Nobes uses the United Kingdom as the benchmark in *Towards 1992.*

The other approach to international accounting is to discuss a particular issue from the perspective of different countries. That is the approach that Nobes uses in this book. *Towards 1992* is a good reference book for those who are interested in particular topics and would like to know how those topics are treated in the different EC countries. Although the book is about the member countries of the EC, it does not treat all countries equally. Most of the focus is on a few key countries, such as the UK, Germany, France, and The Netherlands. However, practices in the other member nations are discussed periodically, and other countries—such as Sweden and Switzerland—are also mentioned.

The first chapter identifies seven major factors that explain the causes of financial reporting differences in the EC: legal systems, providers of finance, taxation, the accountancy profession, inflation, theory, and accidents. It is too bad that Nobes did not identify some of the other major factors that have been referenced in numerous studies and then focus on his seven key factors.

Nobes then concentrates on European differences in financial reporting in a few key areas such as fairness and taxation. Those familiar with *Comparative International Accounting* will recognize the chapter on classification of financial reporting in Europe. It provides an excellent approach to classifying accounting differences worldwide. Nobes adds a different classification scheme of some European countries that differs from, and adds to, his original classification scheme.

The discussion of harmonization and subsequent differences in accounting and reporting processes is very useful. Excerpts from annual reports of companies in different European countries are used to illustrate different accounting practices, such as tangible fixed assets, currency translation, and so on.

There are also three useful appendices: extract from the Annual Report of Total Oil (a French company), extract from the Annual Report of AEG (a German company), and a glossary of some accounting terms in English, Spanish, French, and German. Some of the most useful information contained in the annual reports of European companies is voluntary, and the extracts in the appendices generally contain only the financial statements and footnotes.

In general, the book accomplishes what it sets out to: it is a description of accounting in Europe that is written for practitioners from the perspective of specific accounting topics. However, the book is valuable for the academic that wants to use illustrations of European practices while covering topics in a domestic course.

LEE H. RADEBAUGH
Associate Dean
Marriott School of Management

Brigham Young University

IRWIN WINSTEN, *Accounting Software and the Microcomputer: A Practical Guide to Evaluation and Implementation* (New York: John Wiley & Sons, Inc., 1990, pp. xi, 356, $85.00).

The author's primary objective is to provide "practical" guidance for evaluating and implementing accounting software for the microcomputer. He states in the preface that this guidance is intended for four groups of readers: (1) software users, (2) CPAs, accountants, and business consultants, (3) business educators, and (4) accounting software publishers and designers.

I believe the book is quite unique. It contains a wealth of helpful information, especially for *evaluating* microcomputer accounting software, which I have not seen provided in any other single source. Unfortunately, this information is not presented in a format that makes it easily usable.

Within the first two chapters, the author provides a good discussion of why selecting microcomputer software is so difficult and frustrating. This in itself is a real service for small business personnel who are about to begin software selection, especially in organizations currently using manual accounting methods. However, the primary strength of the book is the author's guidance for evaluating microcomputer accounting software found in chapters five through eight.

In chapters five through eight, the author alerts the reader to an overwhelming number of points and issues that should be considered before a software package is selected. Most of these are specific, concrete issues about the way in which different accounting systems might have been designed and how the resulting designs might function. Most, if not all, of these issues would be familiar to accountants or users of accounting systems, even manual systems. Although some of the issues addressed in these chapters have control implications, the author includes a separate chapter to specifically address control.

Control is probably one of the most important considerations when selecting or designing an accounting system. However, the author's treatment of control in this chapter is not as strong as it could be. This is due, in part, to the rambling narrative style the author uses throughout the book, and, in part, to the fact that many control issues were alluded to in previous chapters and therefore are not thoroughly covered in this chapter. Stronger guidance can be found in a number of accounting systems texts, however, in fairness to the author, these other sources are not written for practitioners and would not generally be sought out by practitioners. In addition, these other sources do not focus specifically on software for the microcomputer.

The author's guidance on installing an accounting system is less concrete and less specific than the guidance he provides for evaluating accounting software. Although this may be unavoidable due to the many organizational characteristics that can differ from one company to another, it makes the guidance less usable and, therefore, less practical.

The greatest deficiency of the book is the style with which the author presents his guidance. The author presents everything in a narrative form that, when used exclusively, doesn't work well because of the large volume of technical and detailed information. The presentation could have been improved if the author had included an appendix that summarized the guidance in outline form and if he had provided structured checklists like those found in CPA firms' internal control questionnaires. In this way, practitioners could more easily see the organization of details. In addition, the checklists could be used during the evaluation and installation processes without referring back to the narrative of the text unless clarification was necessary. Unfortunately, readers of this first edition will have to produce these summaries and checklists on their own to make the guidance usable.

Overall, I believe the book could be very useful to each of its target groups. However, because of its total reliance on a narrative style, I believe readers will not find it as practical as the author intended.

PHILLIP D. HARSHA
Associate Professor of Accounting
Southwest Missouri State University

BOOKS RECEIVED AUGUST 8, 1990 TO NOVEMBER 10, 1990

GRAHAM BANNOCK, *Taxation in the European Community: The Small Business Perspective* (London: Paul Chapman Publishing, Ltd., 1990, pp. ix, 130, $39.95).

WILLIAM T. BAXTER, *Studies in Accounting* (New York: Garland Publishing, Inc., 1989, pp. vi, xii, 455, $85.00).

JOSEPH R. BLASI, *Employee Ownership: Revolution or Ripoff?* (New York: Harper Collins Publishers, 1990, pp. xiv, 334, $18.95, paper).

THOMAS L. BONN, *Heavy Traffic & High Culture* (New York: Meridian, 1989, pp. xii, 240, $10.95, paper).

J. E. BORITZ, *Approaches to Dealing with Risk and Uncertainty* (Toronto: The Canadian Institute of Chartered Accountants, 1990, pp. xxiii, 132, CAD $19.50).

T. E. COOKE, *An Empirical Study of Financial Disclosure by Swedish Companies* (New York: Garland Publishing, Inc., 1989, pp. xii, 381, $85.00).

HUGH M. COOMBS and J. R. EDWARDS, *Accountability of Local Authorities in England and Wales, 1831–1935* (New York: Garland Publishing, Inc., 1990, Volume 1: pp. xvi, 288; Volume 2: pp. ii, 425, $134.00).

PHILLIP G. COTTELL, JR., and TERRY M. PERLIN, *Accounting Ethics: A Practical Guide for Professionals* (Westport, CT: Quorum Books, 1990, pp. xii, 171, $39.95).

PHILIP B. CROSBY, *Let's Talk Quality: 96 Questions You Always Wanted to Ask Phil Crosby* (New York: Plume, pp. x, 208, $9.95, paper).

DAVID A. DAVENPORT, *The Equity Manager Search* (New York: Quorum Books, 1990, pp. xiii, 205, $45.00).

JOHN DONALDSON, *Key Issues in Business Ethics* (London: Academic Press, 1989, pp. xxiv, 227, $19.95, paper).

CHRISTINA S. R. DRUMMOND and ALISTER K. MASON, *Guide to Accounting Pronouncements and Sources*, Second Edition (Toronto: The Canadian Institute of Chartered Accountants, 1990, pp. v, 157, CAD $29.50).

M. J. R. GAFFIKIN, *Accounting Methodology and the Work of R. J. Chambers* (New York: Garland Publishing, Inc., 1989, pp. x, 236, $50.00).

ALAN GART and DAVID J. NYE, *Insurance Company Finance and Investments*, Second Edition (Malvern, PA: Insurance Institute of America, 1990, pp. xvii, 497, $25.00).

CLARKSON GORDON & CO., *The Story of the Firm 1864–1964* (New York: Garland Publishing, Inc., 1989, pp. xvii, 148, $55.00).

O. FINLEY GRAVES, GRAEME W. DEAN, and FRANK L. CLARKE, *Schmalenbach's Dynamic Accounting and Price-Level Adjustments: An Economic Consequences Explanation* (New York: Garland Publishing, Inc., 1989, pp. ix, 166, $42.00).

LEONARD MERVYN HELMORE, *The District Auditor* (New York: Garland Publishing, Inc., 1990, pp. x, 220, $47.00).

ROXANNE THERESE JOHNSON, *An Analysis of the Early Record Keeping in the DuPont Company 1800–1818* (New York: Garland Publishing, Inc., 1989, pp. viii, 105, $38.00).

J. EDWARD KETZ, RAJIB K. DOOGAR, and DAVID E. JENSEN, *A Cross-Industry Analysis of Financial Ratios: Comparabilities and Corporate Performance* (New York: Quorum Books, 1990, pp. x, 218, $49.95).

F. W. LaFRENTZ & CO., *A Half Century of Accounting: The Story of F. W. LaFrentz & Co.* (New York: Garland Publishing, Inc., 1989, pp. 53, $30.00).

T. A. LEE, Editor, *The Closure of the Accounting Profession* (New York: Garland Publishing, Inc., 1990, Volume 1: pp. viii, 369; Volume 2: pp. iv, 300, $118.00).

T. A. LEE and D. P. TWEEDIE, *Shareholder Use and Understanding of Financial Information* (New York: Garland Publishing, Inc., 1990, pp. xxxii, 177, x, 209, $75.00).

DONALD L. MADDEN and JAMES R. HOLMES, *Management Accountants: Responding to Change—An Exploratory Study* (Montvale, NJ: National Association of Accountants, 1990, pp. xvii, 63, $13.95, paper).

WALTER W. MANLEY II with WILLIAM A. SHRODE, *Critical Issues in Business Conduct: Legal, Ethical, and Social Challenges for the 1990s* (Westport, CT: Quorum Books, 1990, pp. xxii, 309, $45.00).

PAUL J. MIRANTI, JR., *Accounting Comes of Age: The Development of an American Profession: 1886–1940* (Chapel Hill, NC: The University of North Carolina Press, 1990, pp. xi, 275, $29.95).

MAURICE MOONITZ, *Selected Writings* (New York: Garland Publishing, Inc., 1990, Volume I: pp. xvi, 263; Volume II: pp. vi, 304, $160.00).

BELVERD E. NEEDLES, JR., and VERNON K. ZIMMERMAN, Editors, *Comparative International Accounting Educational Standards* (Champaign, IL: University of Illinois Center for International Education and Research in Accounting, 1990, pp. iv, 243, $10.00, paper).

PACIFIC RESEARCH PUBLICATIONS, *Database of Accounting Research* (Diamond Bar, CA: 1989, Pacific Research Publications, 9 diskettes, $60.00).

LEE D. PARKER and O. FINLEY GRAVES, Editors, *Methodology and Method in History: A Bibliography* (New York: Garland Publishers, Inc., 1989, pp. xv, 246, $70.00).

ROBERT H. PARKER, Editor, *Accounting in Australia: Historical Essays* (New York: Garland Publishing, Inc., 1990, pp. xii, 559, $145.00).

REVIEW OF THE ACCOUNTING DISCIPLINE IN HIGHER EDUCATION, *Accounting in Higher Education: Volume 1, Main Report and Recommendations* (Canberra: Australian Government Publishing Service, 1990, pp. xxxiv, 237).

MICHAEL J. SANDRETTO, *Reporting, Control, and Analysis of Property, Plant, and Equipment* (Montvale, NJ: National Association of Accountants, 1990, pp. x, 75, $17.95, paper).

ARNOLD L. SCHWARTZ, *Dynamic Professional Selling* (New York: Plume, 1990, pp. xi, 166, $9.95, paper).

LEONARD SPACEK, *The Growth of Arthur Andersen & Co. 1928–1973: An Oral History* (New York: Garland Publishing, Inc., pp. ix, 349, $65.00).

ANDREW W. STARK, Editor, *The Cash Recovery Rate Approach to the Estimation of Economic Performance* (New York: Garland Publishing, Inc., 1990, pp. xii, 153, $55.00).

HANS R. STOLL, Editor, *International Finance & Financial Policy* (New York: Quorum Books, 1990, pp. xxi, 251, $45.00).

THEODOR SWANSON, *Touche Ross: A Biography* (New York: Garland Publishing, Inc., 1990, pp. 96, 9, $45.00).

TERRIE E. TROXEL and GEORGE E. BOUCHIE, *Property-Liability Insurance Accounting and Finance*, Third Edition (Malvern, PA: American Institute for Property and Liability Underwriters, 1990, pp. xv, 352, $20.00).

MILTON F. USRY, Editor, *The Mary Ball Washington Forum Series in Accounting Education: Volume 5* (Pensacola, FL: The University of West Florida, 1990, pp. 70, Free).

INFORMATION FOR AUTHORS

EDITORIAL POLICY

According to the policies set by the Publications Committee (which were endorsed by the Executive Committee and were published in the *Accounting Education News*, June 1987), *The Accounting Review* "should be viewed as the premier journal for publishing articles reporting the results of accounting research and explaining and illustrating related research methodology. The scope of acceptable articles should embrace any research methodology and any accounting-related subject, as long as the articles meet the standards established for publication in the journal.... No special sections should be necessary. The primary, but not exclusive, audience should be—as it is now—academicians, graduate students, and others interested in accounting research."

The primary criterion for publication in *The Accounting Review* is the significance of the contribution an article makes to the literature.

The efficiency and effectiveness of the editorial review process is critically dependent upon the actions of both authors submitting papers and the reviewers. Authors accept the responsibility of preparing research papers at a level suitable for evaluation by independent reviewers. Such preparation, therefore, should include subjecting the manuscript to critique by colleagues and others and revising it accordingly prior to submission. The review process is not to be used as a means of obtaining feedback at early stages of developing the research.

Reviewers and associate editors are responsible for providing critically constructive and prompt evaluations of submitted research papers based on the significance of their contribution and on the rigor of analysis and presentation. Associate Editors also make editorial recommendations to the editor.

MANUSCRIPT PREPARATION AND STYLE

The Accounting Review's manuscript preparation guidelines follows (with a slight modification) the B-format of the *Chicago Manual of Style* (13th ed.; University of Chicago Press). Another helpful guide to usage and style is *The Elements of Style*, by William Strunk, Jr., and E. B. White (Macmillan). Spelling follows *Webster's International Dictionary*.

FORMAT

1. All manuscripts should be typed on one side of 8½ × 11" good quality paper and be double spaced, except for indented quotations.
2. Manuscripts should be as concise as the subject and research method permit, generally not to exceed 7,000 words.
3. Margins of at least one inch from top, bottom, and sides should facilitate editing and duplication.
4. To assure anonymous review, authors should not identify themselves directly or indirectly in their papers. Single authors should not use the editorial "we."
5. A cover page should show the title of the paper, the author's name, title, and affiliation, any acknowledgements, and a footnote indicating whether the author would be willing to share the data (see last paragraph in this statement).

Pagination: *All* pages, including tables, appendices, and references, should be serially numbered. The first section of the paper should be untitled and unnumbered. Major sections may be numbered in roman numerals. Subsections should not be numbered.

Numbers: Spell out numbers from one to ten, except when used in tables and lists, and when used with mathematical, statistical, scientific, or technical units and quantities, such as distances, weights, and measures. For example: *three days*; *3 kilometers*; *30 years*. All other numbers are expressed numerically. Generally when using approximate terms spell out the number, for example, *approximately thirty years*.

Percentages and Decimal Fractions: In nontechnical copy use the word *percent* in the text; in technical copy the symbol % is used. (See the *Chicago Manual* for discussion of these usages.)

Hyphens: Use a hyphen to join unit modifiers or to clarify usage. For example: a *well-presented analysis*; *re-form*. See *Webster's* for correct usage.

Key Words: The synopsis is to be followed by four *key words* that will assist in indexing the paper.

SYNOPSIS

A synopsis of about 400 words should be presented on a separate page immediately preceding the text. The synopsis also serves as the "introduction." The synopsis should be nonmathematical and include a readable summary of the research question, method, and the significance of the findings and contribution. The title, but not the author's name or other identification designations, should appear on the synopsis page.

TABLES AND FIGURES

The author should note the following general requirements:

1. Each table and figure (graphic) should appear on a separate page and should be placed at the end of the text. Each should bear an arabic number and a complete title indicating the exact contents of the table or figure.
2. A reference to each graphic should be made in the text.
3. The author should indicate by marginal notation where each graphic should be inserted in the text.
4. Graphics should be reasonably interpreted without reference to the text.
5. Source lines and notes should be included as necessary.

Equations: Equations should be numbered in parentheses flush with the right-hand margin.

DOCUMENTATION

Citations: Work cited should use the "author-date system" keyed to a list of works in the reference list (see below). Authors should make an effort to include the relevant page numbers in the cited works.

1. In the text, works are cited as follows: authors' last name and date, without comma, in parenthesis: for example, (Jones 1987); with two authors: (Jones and Freeman 1973); with more than two: (Jones et al. 1985); with more than one source cited together (Jones 1987; Freeman 1986); with two or more works by one author: (Jones 1985, 1987).
2. Unless confusion would result, do not use "p." or "pp." before page numbers: for example, (Jones 1987, 115).
3. When the reference list contains more than one work of an author, the suffix *a*, *b*, etc. follows the date in the text citation: for example, (Jones 1987a) or (Jones 1987a; Freeman 1985b).
4. If author's name is mentioned in the text, it needs not be repeated in the citation; for example, "Jones (1987, 115) says. . . ."
5. Citations to institutional works should use acronyms or short titles where practicable; for example, (AAA ASOBAT 1966); (AICPA Cohen Commission Report 1977). Where brief, the full title of an institutional work might be shown in a citation: for example, (ICAEW *The Corporate Report* 1975).
6. If the manuscript refers to statutes, legal treatises, or court cases, citations acceptable in law reviews should be used.

Reference List: Every manuscript must include a list of references containing only those works cited. Each entry should contain all data necessary for unambiguous identification. With the author-date system, use the following format recommended by the *Chicago Manual*:

1. Arrange citations in alphabetical order according to surname of the first author or the name of the institution responsible for the citation.
2. Use author's initials instead of proper names.
3. Dates of publication should be placed immediately after author's name.
4. Titles of journals should not be abbreviated.
5. Multiple works by the same author(s) should be listed in chronological order of publication. Two or more works by the same author(s) in the same year are distinguished by letters after the date.
6. Inclusive page numbers are treated as recommended in *Chicago Manual* section 8.67.

Sample entries are as follows:

American Accounting Association, Committee on Concepts and Standards for External Financial Reports. 1977. *Statement on Accounting Theory and Theory Acceptance*. Sarasota, FL: AAA.

Fabozzi, F., and I. Pollack, eds. 1987. *The Handbook of Fixed Income Securities*. 2d ed. Homewood, IL: Dow Jones-Irwin.

Demski, J. S., and D. E. M. Sappington. 1989. Hierarchical structure and responsibility accounting. *Journal of Accounting Research* 27: 40–58.

Dye, R., B. Balachandran, and R. Magee. 1989. Contingent fees for audit firms. Working paper, Northwestern University, Evanston, IL.

Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, United Kingdom: Cambridge University Press.

Porcano, T. M. 1984a. Distributive justice and tax policy. *The Accounting Review* 59 (October): 619–36.

———. 1984b. The perceived effects of tax policy on corporate investment intentions. *The Journal of the American Taxation Association* (Fall): 7–19.

Shaw, W. H. 1985. Empirical evidence on the market impact of the safe harbor leasing law. Ph.D. dissertation, University of Texas, Austin.

Sherman, T. M., ed. 1984. *Conceptual Framework for Financial Accounting*. Cambridge, MA: Harvard Business School.

Footnotes: Footnotes are not used for documentation. Textual footnotes should be used only for extensions and useful excursions of information that if included in the body of the text might disrupt its continuity. Footnotes should be double spaced and numbered consecutively throughout the manuscript with superscript Arabic numerals. Footnotes are placed at the end of the text.

SUBMISSION OF MANUSCRIPTS

Authors should note the following guidelines for submitting manuscripts:

1. Manuscripts currently under consideration by another journal or other publisher should not be submitted. The author must state that the work is not submitted or published elsewhere.
2. In the case of manuscripts reporting on field surveys or experiments, four copies of the instrument (questionnaire, case, interview plan, or the like) should be submitted.
3. Four copies should be submitted together with a check in U.S. funds for $50.00 for members or $100.00 for nonmembers of the AAA made payable to the American Accounting Association. Effective January 1990, the submission fee is nonrefundable.
4. The author should retain a copy of the paper.
5. Revisions must be submitted within 12 months from request, otherwise they will be considered new submissions.

COMMENTS

Comments on articles previously published in *The Accounting Review* will be reviewed (anonymously) by two reviewers in sequence. The first reviewer will be the author of the original article being subjected to critique. If substance permits, a suitably revised comment will be sent to a second reviewer to determine its publishability in *The Accounting Review*. If a comment is accepted for publication, the original author will be invited to reply. All other editorial requirements, as enumerated above, also apply to proposed comments.

POLICY ON REPRODUCTION

An objective of *The Accounting Review* is to promote the wide dissemination of the results of systematic scholarly inquiries into the broad field of accounting.

Permission is hereby granted to reproduce any of the contents of the *Review* for use in courses of instruction, as long as the source and American Accounting Association copyright are indicated in any such reproductions.

Written application must be made to the Editor for permission to reproduce any of the contents of the *Review* for use in other than courses of instruction—e.g., inclusion in books of readings or in any other publications intended for general distribution. In consideration for the grant of permission by the *Review* in such instances, the applicant must notify the author(s) in writing of the intended use to be made of each reproduction. Normally, the *Review* will not assess a charge for the waiver of copyright.

Except where otherwise noted in articles, the copyright interest has been transferred to the American Accounting Association. Where the author(s) has (have) not transferred the copyright to the Association, applicants must seek permission to reproduce (for all purposes) directly from the author(s).

POLICY ON DATA AVAILABILITY

The following policy has been adopted by the Executive Committee in its April 1989 meeting.

"An objective of (*The Accounting Review, Accounting Horizons, Issues in Accounting Education*) is to provide the widest possible dissemination of knowledge based on systematic scholarly inquiries into accounting as a field of professional, research, and educational activity. As part of this process, authors are encouraged to make their data available for use by others in extending or replicating results reported in their articles. Authors of articles which report data dependent results should footnote the status of data availability and, when pertinent, this should be accompanied by information on how the data may be obtained."

APRIL 1991 PLACEMENT ADS

The deadline for information to be included in the "Placement Information" section is two months prior to the date of publication. The REVIEW *is published on the first day of January, April, July, and October. Placement advertising and replies should be sent to the American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233. ATTENTION: ADVERTISING DEPT. All replies to box numbers should be on separate sheets to facilitate forwarding. There is no charge for this service to AAA members and associate members seeking employment. Others will be charged a fee of $150 per insertion. ADS SHOULD NOT EXCEED 150 WORDS.*

#201 THE UNIVERSITY OF MARYLAND European and Asian Divisions seeks excellent teachers for undergraduate courses on U.S. military bases overseas. Minimum requirements include a master's in accounting, one year of recent college teaching experience, and U.S. citizenship. Renewable annual appointments begin August 1991. Excellent for a sabbatical or leave of absence. Benefits include transportation and military base privileges. Frequent travel and the cost of schooling make these positions difficult for those with children. Send resumé to: Dr. Ralph E. Millis, The University of Maryland University College, College Park, MD 20742-1642. AA/EEO.

#202 THE UNIVERSITY OF NEW ORLEANS invites applications for positions in the areas of Auditing, Systems, and Taxation. Rank is at the Assistant or Associate level depending upon qualifications. Candidates should have the Ph.D. (accounting major). To qualify at the Associate rank, individuals must have an established record of scholarly research and a commitment to quality teaching. Salary is negotiable and competitive. The University offers the M.S. in Accounting and the M.S. in Accounting (tax option), as well as the M.B.A. Approximately 550 undergraduate accounting majors are enrolled. The Department has separate AACSB accounting accreditation at both the graduate and undergraduate levels. Candidates should send a resumé to: Professor Philip J. Harmelink, Chairman, Department of Accounting, College of Business Administration, University of New Orleans, LA 70148. UNO is a member of the Louisiana State University System and an Affirmative Action, Equal Opportunity Employer.

#203 THE UNIVERSITY OF CALGARY Faculty of Management invites applications for positions at all levels in the areas of taxation, auditing, financial, and managerial accounting. Candidates should have a PhD or DBA with a major in accounting. ABD's in advanced dissertation stage will be considered. Research and publication requirements will be commensurate with rank upon appointment. Salary is competitive and computer assistance and research support are available. Starting date will be determined by mutual agreement. The University of Calgary has an Employment Equity Program and encourages applications from all qualified candidates, including women, aboriginal people, visible minorities, and people with disabilities. Interested individuals are asked to mail applications to: Murray Davis, PhD. CA, Chair, Accounting Faculty of Management, The University of Calgary, 2500 University Drive, N.W., Calgary, Alberta T2N 1N4.

#204 EASTERN MICHIGAN UNIVERSITY seeks applicants for tenure-track faculty position available Fall, 1991. All specializations considered, with particular interest in information systems. Rank can range from instructor through full professor, depending on qualifications. PhD or DBA with accounting major preferred, advanced doctoral candidates considered. Master's Degree plus professional certification considered for instructor level only. Accounting and/or teaching experience at a collegiate school of business desired. Early reply requested, final deadline June 30, 1991. Submit detailed resumé to: Position FAAA91001, EASTERN MICHIGAN UNIVERSITY, 310 King Hall, Ypsilanti, MI 48197. Phone inquiries to: (313) 487-3320. AA/EOE, women and minorities encouraged to consider these opportunities.

#205 DBA (Accounting), CPA, CMA seeks tenure-track position. My experience includes 14 years teaching a variety of undergraduate business and accounting courses. I have eight years experience in private accounting. I am actively involved in my community and college. Teaching preferences include cost/managerial accounting and governmental accounting. I prefer a small to medium school located in the South.

#206 TAX POSITION: Tax partner in 45-attorney law firm desires tenure track or visiting position at AACSB-accredited school. J.D./LL.M. (Taxation)/CPA, with strong academic record (top five percent in both undergraduate accounting and LL.M. in Taxation programs). More than 20 published articles. Frequent lecturer to attorneys and CPAs on tax matters. Former adjunct professor who meets AACSB faculty standards for (1) doctorate, and (2) professional certification (while teaching tax subjects). Big Six experience. Worked in IRS Appeals Division while earning LL.M. Broad tax background and experience in tax litigation would add important diversity to faculty. Wife earns Ph.D. in Spring, 1991 and desires faculty position in forest social sciences or landscape architecture. For resumé, write: Tax Position, 1824 North 19th Street, Boise, Idaho 83702-0707. Telephone (208) 336-6811.

#207 CLEVELAND STATE UNIVERSITY seeks candidates for faculty positions in accounting. Rank dependent upon qualifications. Candidates for assistant professor should have Ph.D. or D.B.A. in Accounting, or at least be in advanced dissertation stage. Professional certification (CPA, CMA, CISA, or CIA) and research capacity favorably considered. Candidates for associate and full professor must have doctorate and record of publications. Salary is negotiable. Send resumé to: Dr. Elise G. Jancura, Chairperson, Departments of Accounting and Business Law, Cleveland State University, E. 24th & Euclid Avenue, Cleveland, OH 44115. Equal Opportunity Employer, m/f/h.

#208 UNIVERSITY OF UTAH SCHOOL OF ACCOUNTING invites applications for Associate or Full Professor of Accounting. Applicants should possess a Ph.D. and have significant research achievements. In addition to teaching responsibilities, all applicants are expected to perform scholarly research. Applicants with primary interest in financial accounting or information systems are urged to apply. Salaries are competitive and fringe benefits are excellent. Programs in accounting are offered at the bachelor, master, and Ph.D. levels. Application deadline is May 1, 1991, or until position is filled. Please send application and vita to: Dr. Ron N. Bagley, Director, School of Accounting, Graduate School of Business, University of Utah, Salt Lake City, Utah 84112. The University of Utah is an Equal Opportunity Affirmative Action Employer and encourages nominations and applications from women and minorities.

#209 EASTERN ILLINOIS UNIVERSITY invites applications for Accountancy Faculty. Specialty in Accounting Information Systems preferred, other areas of emphasis will be considered. Teaching and business experience preferred, beginning Fall, 1991. Doctorate required. Very competitive salary; exempt from social security; retirement contributions tax sheltered; life and health insurance provided; several TSA plans available. Opportunities for consulting and summer research; Business Development Center; Small Business Institute. The Lumpkin College of Business is endowed by Lumpkin Foundation and has 2,539 undergraduates in seven programs offered through four departments, also 100 MBA students. Recently moved into new building. EIU is a regional residential institution with a tradition of excellent instruction, high quality student body of 10,000 located in east central Illinois. Community population of 20,000 with easy access to Chicago, St. Louis, and Indianapolis. Contact: Dr. Frank L. Clark, 232 Lumpkin Hall, EIU, Charleston, IL 61920. (217-581-3023). AA/EOE.

#210 TEMPORARY POSITION SOUGHT. Experienced administrator/faculty member seeks appropriate position for all or part of the 1991-1992 academic year while on sabbatical. Presently Chair of a Department of Accounting that is fully AACSB-accredited at A and C levels.

#211 ANTICIPATED POSITION IN BUSINESS ADMINISTRATION. Date of appointment: August 26, 1991. Salary: Commensurate with qualifications and experience. Job description: Teaching 12 semester credits of business courses, including accounting systems, quantitative management, and other accounting courses. Tenure track or academic staff. Qualifications required: Ph.D. preferred or MBA or appropriate master's degree in business. Business experience desirable. Direct inquiries to: Don C. Aabel, Chair, Search and Screen Committee, Department of Business Administration, University of Wisconsin-River Falls, River Falls, WI 54022 (715) 425-3335. Deadline for receipt of application materials: Until filled. UW-RF is committed to achieving diversity in its academic community. Women and minorities are strongly encouraged to apply.

#212 HENDRIX COLLEGE, a coeducational, residential college, invites applications for a tenure-track position in accounting which will open September, 1991. Teaching responsibilities will include introductory and advanced financial accounting and auditing. The terminal degree for rank of assistant professor is an MBA with a CPA or CMA certificate. Relevant work experience and computer skills are highly desirable. A strong commitment to teaching is expected. Hendrix offers a competitive salary and liberal fringe benefits compensation package. The deadline for applications is April 20, 1991. Send a resumé, three letters of recommendation, and copies of your transcripts to: Stephen W. Kerr, Department of Economics & Business, Hendrix College, Conway, AR 72032. Women and minorities urged to apply.

#213 Ph.D., CPA with over ten years teaching experience at undergraduate and graduate level. Visiting or tenure-track position desired. Available after June, 1991.

#214 SIMON FRASER UNIVERSITY, Faculty of Business Administration invites applications for tenure-track positions in Managerial Accounting, Financial Accounting, Audit, and Tax Accounting. Rank open. Successful candidate will have strong research and teaching abilities for both undergraduate and graduate courses and Ph.D. in hand or near completion. Candidates eligible for employment in Canada at the time of application are preferred. Simon Fraser University is committed to the principle of equity in employment and offers equal employment opportunities to qualified applicants. This position is subject to budgetary approval. Send curriculum vitae and the names of three referees to: Dr. Bertram Schoner, Chairman, Appointments Committee, Faculty of Business Administration, Simon Fraser University Burnaby, British Columbia V5A 1S6.

#215 POSITION WANTED—Give your students a real world perspective! Practitioner with 28 years real world experience seeks full-time academic position. MBA, CPA. Former National Auditing Partner with international CPA firm; specialist in audit sampling and audit methodology. Numerous publications, including five articles in *Journal of Accountancy*. Presenter and discussant at academic conferences. Developer and instructor of Continuing Professional Education courses. Chairman or member of many professional committees. Please contact: Abraham D. Akresh, 1209 Weymouth Road, Wynnewood, PA 19096 or call 215-642-1742.

#216 LOYOLA COLLEGE IN MARYLAND anticipates a visiting position open for one, possibly two, years, beginning Fall, 1992. We are recruiting an individual to teach primarily in the financial accounting area. Minimum qualifications are a master's degree with professional certification; doctoral degree desirable. Major teaching responsibility at undergraduate level; graduate assignments available dependent on academic credentials. Excellence in teaching is a primary emphasis; professional involvement and research is expected and supported. Salary is competitive. Loyola's business and accounting programs have AACSB accreditation. Please send resumé to: William E. Blouch, Chairman, Department of Accounting, Loyola College in Maryland, 4501 N. Charles Street, Baltimore, MD 21210. Opportunities for an interview are available in Nashville at August AAA National Meeting. EEO/AAE.

#217 THE UNIVERSITY OF MANITOBA is seeking applicants for tenure-track positions in accounting for July 1991. Rank and salary open, based on qualifications and experience. A visiting appointment is also possible. Ph.D. or DBA, completed or near completion, is required. Duties include research teaching at the undergraduate and graduate levels. Salary is competitive and will depend on qualifications, experience, and research record. Winnipeg, a multicultural city of 600,000 offers the best of an urban environment in the heart of Canada's vacation land. The University of Manitoba encourages applications from qualified women and men, including members of visible minorities, aboriginal people, and persons with disabilities. The University offers a smoke-free environment, save for specially designated areas. Priority consideration will be given to Canadian citizens and permanent residents. Applications should be sent to: DR. LAWRENCE I. GOULD, HEAD, DEPARTMENT OF ACCOUNTING AND FINANCE, UNIVERSITY OF MANITOBA, WINNIPEG, MANITOBA, R3T 2N2.

#218 FLORIDA INTERNATIONAL UNIVERSITY invites applications for Director of the School of Accounting. Candidates should have an earned doctorate, professional certification, a strong commitment to teaching excellence, and a history of scholarly activities. Administrative and professional experience is desirable. The School has 28 full-time faculty and offers programs at the bachelor, master, and Ph.D. levels. Appointment will be effective for the 1991 Fall semester. Send letters of application to: Dr. Harvey S. Hendrickson, Chair, Director Search Committee, School of Accounting, Florida International University, Miami, Florida 33199. Florida International University is an Equal Opportunity/Affirmative Action Employer.

#219 CATAWBA COLLEGE invites applications for a tenure-track position in Accounting at the Assistant or Associate Professor level beginning Fall, 1991. Ph.D. and CPA important. Business experience, demonstrated competency in teaching, an interest in participating in professional activities, executive educational programs, and some interest in research are also desirable. Applicant must have competence and be willing to teach in two of the following areas: financial, managerial/cost, and tax. Catawba College is a Liberal Arts institution located in Salisbury, North Carolina, 40 miles north of Charlotte. The Ketner School of Business is housed in a new four and a half million dollar building with up-to-date computer and classroom facilities. Application with a resumé listing three references should be sent to: Dr. Junius H. Terrell, Dean, Ketner School of Business, Catawba College, Salisbury, NC 28144. Application deadline: April 15, 1991, or until filled. An equal opportunity, affirmative action employer.

#220 MASSACHUSETTS INSTITUTE OF TECHNOLOGY Sloan School Accounting Group is seeking new faculty at the junior level effective July 1, 1991. Applicants will be judged on their potential for excellence in research and teaching, although preference will be given to those who can submit completed research. Applicants should submit a current vitae and evidence of research such as publications, working papers, or dissertation proposal and chapters to: Professor Paul Healy, Chairperson of Search Committee, MIT Sloan School of Management, 50 Memorial Drive, Bldg. E52-325H, Cambridge, MA 02139-4307. MIT is an equal opportunity employer and encourages applications from women and minorities.

#221 THE UNIVERSITY OF SAN DIEGO, School of Business, is seeking applications for tenure-track appointments at the Assistant or Associate Professor levels. Visiting positions are also available. The University numbers 5,000 students and the School of Business is AACSB-accredited. Ph.D. required, ABD considered. If you think that a small, quality-oriented school where your input counts, is where you would like to be, please forward your resumé. We will also be talking to potential faculty members at the national AAA meeting in Nashville. Please send your resumé to: Dr. Carol Bruton, School of Business, University of San Diego, San Diego, CA 92110. (619) 260-4854.

PUBLICATIONS
from
The American Accounting Association

RESEARCH SERIES

	AAA Members	Non-Members
Volume No. 1. Investment Analysis and General Price-Level Adjustments By T. R. Dyckman. Published 1969, 76 pages.	$4.00	$6.00
Volume No. 2. Accounting and Information Theory By Baruch Lev. Published 1969, 84 pages.	$4.00	$6.00
Volume No. 3. The Allocation Problem in Financial Accounting Theory By Arthur L. Thomas. Published 1969, 122 pages.	$4.00	$6.00
Volume No. 4. Accounting Controls and the Soviet Economic Reforms of 1966 By Bertrand Horwitz. Published 1970, 74 pages.	$4.00	$6.00
Volume No. 5. Information Evaluation By Gerald A. Feltham. Published 1972, 149 pages.	$4.00	$6.00
Volume No. 6. A Statement of Basic Auditing Concepts By the Committee on Basic Auditing Concepts. Published 1973, 58 pages.	$4.00	$6.00
Volume No. 7. Valuation of Used Capital Assets By Carl R. Beidleman. Published 1973, 84 pages.	$4.00	$6.00
Volume No. 8. Obtaining Agreement on Standards in the Accounting Profession By Maurice Moonitz. Published 1974, 93 pages.	$4.00	$6.00
Volume No. 9. The Allocation Problem: Part Two By Arthur L. Thomas. Published 1974, 194 pages.	$4.00	$6.00
Volume No. 10. Theory of Accounting Measurement By Yuji Ijiri. Published 1975, 210 pages.	$4.00	$6.00
Volume No. 11. The Impact of Audit Frequency on the Quality of Internal Control By Russell M. Barefield. Published 1975, 86 pages.	$4.00	$6.00
Volume No. 12. A Comparative Analysis of Selected Income Measurement Theories in Financial Accounting By James A. Anderson. Published 1976, 120 pages.	$4.00	$6.00
Volume No. 13. Measurement and Accounting Information Criteria By Theodore Jave Mock. Published 1976, 116 pages.	$4.00	$6.00
Volume No. 14. Phantasmagoric Accounting By Robert E. Jensen. Published 1976, 209 pages.	$4.00	$6.00
Volume No. 15. Adaptive Behavior: Management Control and Information Analysis By Hiroyuki Itami. Published 1977, 164 pages.	$4.00	$6.00
Volume No. 16. Earnings or Cash Flows: An Experiment on Functional Fixation and the Valuation of the Firm By Rashad Abdel-khalik and Thomas F. Keller. Published 1979, 101 pages.	$4.00	$6.00
Volume No. 17. Human Information Processing in Accounting By Robert H. Ashton. Published 1982, 215 pages.	$6.00	$9.00
Volume No. 18. Triple-Entry Bookkeeping and Income Momentum By Yuji Ijiri. Published 1982, 53 pages.	$6.00	$9.00
Volume No. 19. Review of Forecasts: Scaling and Analysis of Expert Judgments Regarding Cross-Impacts of Assumptions on Business Forecasts and Accounting Measures By Robert E. Jensen. Published 1983, 235 pages.	$11.00	$15.00
Volume No. 20. A Contract Perspective on Accounting Valuation By John S. Hughes. Published 1984, 92 pages.	$7.00	$10.00
Volume No. 21. Security Analyst Multi-Year Earnings Forecasts and the Capital Market By Philip Brown, George Foster and Eric Noreen. Published 1985, 172 pages.	$8.00	$12.00
Volume No. 22. Essays in Accounting Theory By Carl Thomas Devine, in five volumes. Published 1985. Hardcover Set (all five volumes) in two books, total 892 pages.	$50.00	$80.00
Paperback single volumes: Volume I, 172 pages Volume II, 160 pages Volume III, 144 pages Volume IV, 192 pages Volume V, 224 pages	$8.00 each	$12.00 each

	AAA Members	Non-Members
Volume No. 23. Behavior of Test Statistics in the Auditing Environment By Gordon L. Duke, Robert A. Leitch, and John Neter. Published 1986, 178 pages.	$8.00	$12.00
Volume No. 24. Probabilistic Choice Models Information By Ronald W. Hilton. Published 1985, 194 pages.	$11.00	$15.00
Volume No. 25. Univariate Time-Series Analysis of Quarterly Earnings: Some Unresolved Issues By William S. Hopwood and James C. McKeown. Published 1986, 52 pages	$7.00	$10.00
Volume No. 26. Comparison of Audit Methodologies of Large Accounting Firms. By Barry E. Cushing and James K. Loebbecke. Published 1986, 97 pages.	$7.00	$10.00
Volume No. 27. The Effect of Local Government Accounting Practices on Accounting Numbers and Creditor Decisions By Robert W. Ingram and Ronald M. Copeland. Published 1986, 110 pages.	$7.00	$10.00
Volume No. 28. Methodological Foundations of Standardsetting for Corporate Financial Reporting By James C. Gaa. Published 1988, 224 pages.	$13.00	$17.00
Volume No. 29. The Marginal Approach to Joint Cost Allocation: Theory and Application By Rene P. Manes and Agnes C. S. Cheng. Published 1988, 219 pages.	$16.00	$20.00
Volume No. 30. Analysis of Financial Statements: Financial Accounting and the Capital Market By Nicholas J. Gonedes and Nicholas Dopuch. Published 1988, 421 pages.	$18.00	$25.00
Volume No. 31. Momentum Accounting and Triple-Entry Bookkeeping By Yuji Ijiri. Published 1989, 151 pages.	$8.00	$12.00

ACCOUNTING EDUCATION SERIES

Volume No. 2. Researching the Accounting Curriculum: Strategies for Change By William Ferrara. Published 1975, 227 pages.	$4.00	$6.00
Volume No. 3. The Core of the Curriculum for Accounting Majors By Richard E. Flaherty. Published 1979, 234 pages.	$6.00	$8.00
Volume No. 4. Empirical Research in Accounting: A Methodological Viewpoint By A. Rashad Abdel-khalik and Bipin B. Ajinkya. Published 1979, 125 pages.	$4.00	$6.00
Volume No. 5. Developing Communications Skills for the Accounting Profession By Robert W. Ingram and Charles R. Frazier. Published 1980, 66 pages.	$3.00	$6.00
Volume No. 6. Accounting Structured in APL By Yuji Ijiri. Published 1984, 151 pages.	$6.00	$8.00
Volume No. 7. Computer Assisted Analytical Review System By William R. Kinney, Jr., Gerald L. Salamon, and Wilfred C. Uecker. Published 1986, 76 pages.	$8.00	$12.00
Volume No. 8. Computer Integration into the Accounting Curriculum: Case Studies Edited by Robert W. Ingram. Published 1987, 248 pages.	$5.00	$10.00
Volume No. 9. A Framework for the Development of Accounting Education Research By Jan R. Williams, Hartwell Herring, Mikel Tiller, and James Scheiner. Published 1988, 189 pages.	$6.00	$10.00
User's Guide and Diskettes	$30.00	$40.00
Volume No. 10. Reorienting Accounting Education: Reports on the Environment, Professoriate, and Curriculum of Accounting By Joseph J. Schultz, Jr. Published 1989, 212 pages.	$13.00	$17.00

MISCELLANEOUS PUBLICATIONS

Monograph No. 3. An Introduction to Corporate Accounting Standards By W. A. Paton and A. C. Littleton. Published 1940, 156 pages.	$3.00	$6.00
Monograph No. 5. Structure of Accounting Theory By A. C. Littleton. Published 1953, 234 pages.	$3.00	$6.00
Monograph No. 6. The Philosophy of Auditing By R. K. Mautz and Hussein A. Sharaf. Published 1961, 248 pages.	$3.00	$6.00
Price Level Changes and Financial Statements, Case Studies of Four Companies By Ralph C. Jones. Published 1955, 179 pages.	$3.00	$3.00
Price Level Changes and Financial Statements, Basic Concepts and Methods By Perry Mason. Published 1956, 32 pages.	$1.00	$1.00
Effects of Price Level Changes on Business Income, Capital and Taxes By Ralph C. Jones. Published 1956, 199 pages.	$3.25	$3.25

	AAA Members	Non-Members
Accounting & Reporting Standards for Corporate Financial Statements & Preceding Statements & Supplements Published 1957, 64 pages.	$0.50	$1.00
A Statement of Basic Accounting Theory By The Committee to Prepare a Statement of Basic Accounting Theory. Published 1966, 100 pages.	$3.00	$6.00
A Statement of Accounting Principles By Thomas H. Sanders, Henry R. Hatfield, Underhill Moore. Published 1938, 138 pages.	$3.00	$6.00
Statement on Accounting Theory and Theory Acceptance By The Committee on Concepts and Standards for External Financial Reports. Published 1977, 61 pages.	$3.00	$6.00
Accounting Students and Abstract Reasoning By George E. Shute. Published 1979, 71 pages.	$3.00	$6.00
A Review of Innovative Approaches to College Teaching By Beatrice Gross, Ronald Gross. Published 1980, 68 pages.	$3.00	$6.00
Notable Contributions to the Periodical International Accounting Literature—1975-78 By The International Accounting Section of AAA. Published 1979, 168 pages.	$6.00	$8.00
Publication Opportunities for Tax Researchers By Jane O. Burns, Janis R. Reeder, and James L. Wittenbach. Published 1988, 241 pages.	$3.00	$3.00
Comparative International Auditing Standards Edited by Belverd E. Needles, Jr. Published 1985, 193 pages.	$10.00	$12.00
Comparative International Taxation Edited by Kathleen E. Sinning. Published 1986, 86 pages.	$5.00	$5.00
A Compendium of Research on Information and Accounting for Managerial Decision and Control in Japan By Seiichi Sato, Kyosuke Sakate, Gerhard G. Mueller, and Lee H. Radebaugh. Published 1982, 159 pages.	$8.00	$10.00
Annotated International Accounting Bibliography—1972-1981 By Abdel M. Agami and Felix P. Kollaritsch. Published 1983, 175 pages.	$5.00	$7.00
Tax Planning for Educators By American Taxation Association. Published 1987, 90 pages	$6.00	$8.00
Trueblood—Auditing & Case Studies By Touche Ross Foundation & American Accounting Association. Updated 1988. Set—2 books	$10.00	$10.00
Proceedings of the V International Accounting Congress on Accounting Education Edited by Juan M. Rivera and Konrad W. Kubin. Published by International Section, 1987, 100 pages.	$8.00	$10.00
Cases in International Accounting Edited by James A. Schweikart. Published by International Section, 1987, 142 pages.	$10.00	$14.00
Cost Accounting, Robotics, and The New Manufacturing Environment Edited by Robert Capettini and Donald K. Clancy. Vanderbilt Symposium 1987. Published by Management Accounting Section, 1987, 198 pages.	$5.00	$5.00
Research Opportunities in Auditing: The Second Decade A. Rashad Abdel-khalik and Ira Solomon, Editors. Published by Auditing Section, 1989, 216 pages.	$10.00	$15.00
An Index to the Accounting Review 1926-1978 By Gary John Previts and Bruce Committee, 641 pages.	$10.00	$20.00

All members of the American Accounting Association are entitled to purchase publications listed at the member prices. Make checks or money orders payable to American Accounting Association. Orders for addresses within the United States will be mailed fourth class, postage paid. Orders going to destinations outside the United States and Canada will be sent by surface rate. In all cases, if special handling is requested, the costs will be assumed by the purchaser.

* Monographs Number 1, 2, 4, and 7 are out of print.